

多特性多方法データを用いて測定の信頼性と妥当性を
定量的に評価するための方法に関する研究

久保 沙織

目次

第1章	問題および目的	1
1.1	信頼性と妥当性	1
1.2	多特性多方法行列と収束的妥当性・弁別的妥当性	6
1.3	多特性多方法行列のための確認的因子分析モデル	7
1.4	問題	18
1.5	研究目的および本稿の構成	21
第2章	研究 I: 信頼性と妥当性の解釈が一意に定まるモデル	23
2.1	問題と目的	23
2.2	モデル	26
2.3	適用例	32
2.4	研究 I 総合考察	42
第3章	研究 II: 特性ごとの合計得点における信頼性と妥当性の検討	44
3.1	問題と目的	45
3.2	モデル	48
3.3	適用例	50
3.4	研究 II 総合考察	61
第4章	ベイズ統計学とマルコフ連鎖モンテカルロ法	63
4.1	ベイズ統計学	63
4.2	マルコフ連鎖モンテカルロ法	65
4.3	ハミルトニアンモンテカルロ法	74
4.4	収束判定	78
第5章	研究 III: 同一立場内に複数の評価者がいる多特性多評価者データ における信頼性と妥当性の検討	83
5.1	問題と目的	83
5.2	モデル	85
5.3	適用例	89
5.4	研究 III 総合考察	94
第6章	総合考察	96

付録	100
引用文献	126

第1章 問題および目的

第1章においては、第2章以降に展開される研究内容に関連した学術的知識についての記述および論文全体を通しての目的の記述がなされる。まず最初に、妥当性研究の歴史を概観する。続いて、多特性多方法データを用いて収束的妥当性・弁別的妥当性を検証するための代表的ないくつかの統計的手法について概説する。最後に、各手法の特徴や課題を踏まえた上で、本論文の目的と構成を述べる。

1.1 信頼性と妥当性

心理学の研究では、研究対象である心理現象を直接的に測定することができないために、その心理現象を反映していると考えられる何らかの物理量を測定することによって、間接的に心理量の測定を行っている(森, 2009)。ここでいう心理現象とは、心理学において興味の対象となる人格特性や学力、行動傾向等を指しており、これらは構成概念(construct)と呼ばれる。直接観測することのできない構成概念は、質問紙調査や検査、テストなどを通して間接的に測定される。

一般に、心理学的測定は信頼性と妥当性という2つの観点から評価される。信頼性とは、測定の安定性のことであり、測定誤差が小さく得点が一貫している程度として定義される(南風原, 2011)。妥当性とは、興味の対象となる構成概念を正しく測定できている程度のことである。とりわけ、構成概念という目に見えないものを測定の対象とすることが多い心理学においては、妥当性の検討が重要な問題となる(森, 吉田, 1990)。しかしながら、妥当性研究の歴史で後述するように、妥当性はその概念を集約的に解釈することが難しく、様々な側面からの検討を要する。そのため、妥当性検証の手続きや、妥当性係数といっ

た指標について、常套の数理的・統計的なアプローチは存在しないというのが現状である。

その一方で、信頼性に関しては数理的に扱いやすいため、様々な統計手法が提案され、実践場面でも積極的に利用されている。信頼性係数は、古典的テスト理論に基づき、“観測変数全体の分散に占める真の得点の分散”と明確に定義される。信頼性係数を実際に推定するための方法として、再検査法、平行検査法、折半法、内的整合性による信頼性係数などが挙げられる。内的整合性とは、ある構成概念を測定する尺度が複数の項目から構成されているときに、それらの項目がすべて一貫して同じ構成概念を測定している程度を表す。心理学研究においてよく用いられるクロンバックの α 係数 (Cronbach, 1951) は、内的整合性による信頼性係数の1つである。また、内的整合性を確認するために因子分析の立場から提案された指標としては、 ω 係数 (McDonald, 1978) がある。

さらに、分散分析の手法を導入することで古典的テスト理論の枠組みを拡張し、測定の信頼性を検討するための手法として、一般化可能性理論 (generalizability theory; Brennan, 1992; Brennan, 2001; Cronbach, Gleser, Nanda & Rajaratnam, 1972; Shavelson & Webb, 1991) がある。古典的テスト理論の枠組みでは、測定値を真値と測定誤差とに分解し、測定値の分散に占める真値の分散の割合によって信頼性を定義している。しかしながら、現実場面では、測定を意図している特性の真値と、偶然的な測定誤差の他にも、測定値の変動に影響を与える要因は複数存在する。古典的テスト理論では、測定誤差は単一で全体的な存在であるとみなされるのに対して、一般化可能性理論では、複数の要因から生じる誤差を別々の存在として扱う (Feldt & Brennan, 1989) ため、これらの要因1つ1つが測定値全体の変動に与える影響を分離して捉えることができる。この際、実験計画法の考え方に基づき、分散分析の手法を適用することで、それぞれの要因がもつ分散成分の影響力を推定する (池田, 1992a)。

妥当性研究の歴史

以下では、妥当性研究の歴史を概観する。妥当性の古典的な定義として、最もよく知られているのは、Kelley (1927) によるものであろう。すなわち、“妥当

性とは、テストが測定していると主張しているものを本当に測定できているかどうかである (The problem of validity is that of whether a test really measures what it purports to measure)”という定義である。ただし、この時点では、“テストが測定していると主張しているもの”が、明確に構成概念に対応していたわけではない。例えば、大学入試における入学後の成績や、採用試験における採用後の成績のように、可能であれば直接測定したいがその時点では実際に測定できない対象が、“テストが測定していると主張しているもの”に相当する。これを基準得点と呼び、多くの場合、妥当性はテスト得点と基準得点との相関によって検証されていた。

1950年以降、妥当性には4つの種類があるとされた。予測的妥当性 (predictive validity)、併存的妥当性 (concurrent validity)、内容妥当性 (content validity)、そして構成概念妥当性 (construct validity) である (APA, 1954; Cronbach & Meehl, 1955)。これに対して、1966年の Standards for Educational and Psychological Tests and Manuals (APA, 1966; 以下、スタンダードと呼ぶ) では、予測的妥当性と併存的妥当性が基準関連妥当性 (criterion-related validity) に合併してまとめられ、大きく3つの種類に分けられた。

基準関連妥当性は、尺度得点が、同じ構成概念を測る他の変数や予測したい変数とどの程度関連しているかを表し、内容妥当性は、測ろうとする領域全体と尺度の内容とがどの程度適切に関連しているかを表し、構成概念妥当性は尺度得点の高低を尺度が測ろうとする構成概念の強弱として解釈できるかどうかを表す (平井, 2006)。基準関連妥当性のうち、予測的妥当性は、個人の将来の基準変数の水準をそれ以前のテスト成績からどの程度予測できるかによって表され、併存的妥当性は、個人の現在の基準に対する位置をテスト得点がどの程度推定できるかによって表される (池田, 1992b)。

上述の3種類の妥当性は並列的に位置づけられ、妥当性の検証にあたっては、どれか1つ以上の妥当性を示せばよいとされてきた (Angoff, 1988)。ただし、1966年のスタンダードでは、これら3つの妥当性の側面は概念的に独立しているというだけであって、ある特定の場面でその中の1つだけが重要であるということとはほとんどなく、テストについての完全な研究には通常これらすべてのタイプの妥当性についての情報が関連している、と述べられている (Messick,

1989)。

1970年代に入ると、妥当性の概念の統一化に向けた動きが加速した。Cronbach (1971) は、妥当性を3つの種類に分類するという流れを引き継いでいるものの、どれか1つの種類のみを検証すればよいというわけではない、ということ強調している。1974年版のスタンダード (APA, 1974) もまた、3種類の区別を存続しつつも、それらはいくまでも便宜的な区別であり、互いに操作的あるいは理論的に関連した妥当性の側面であると述べている。また、当初妥当性の焦点は、ある特定の基準に対する予測の正確さにあったのに対して、この頃から、測定あるいはテスト得点の意味や解釈に焦点が当てられるようになった。Cronbach (1971) は、妥当化されるのはテストではなく、特定の手続きから得られるデータの解釈である、ということを確認している。

そして1985年のスタンダード (APA, 1985) では、妥当性はテスト得点からなされる特定の推測の適切性、有意性、有用性を示す単一の概念であると明言された。こうして現在では、妥当性はいくつかの種類にわけられるものではなく、1つの統合体であると認識されるようになった。Messick (1995) によると、妥当性とは、テストその他のアセスメントによる得点に基づいた推論や行為の十分さと適切性について、それを支える実証的証拠や理論的根拠がどの程度あるかに関する、総合的な評価的判断である。

妥当性を検証するためには、測定が妥当であるとしたら、具体的にどのような条件が満たされるべきか、という妥当性のための必要条件をリストアップし、それらの条件が実際に満たされているかどうかを確認するという手続きをとることとなる (南風原, 2002)。妥当性の証拠を集める側面として、Messick (1995) は、内容的側面、本質的側面、構造的側面、一般化可能性の側面、外的側面、結果的側面の6つを挙げている。以下は、平井 (2006) からの引用である。

1. 内容的側面からの証拠 測定したい構成概念に含まれる要素、含まれない要素を明確に線引きした領域を定義し、尺度内容が領域に対応しているか、十分に代表しているかを示す証拠である。伝統的には、専門家や尺度のユーザー、被験者の代表などの判断によって行われる。

2. 本質的側面からの証拠 尺度項目への反応プロセスが心理学的に説明できるという証拠である。理論的な証拠としては、反応プロセスモデルがあり、実証的な証拠としては think aloud プロトコル、被験者の回答や反応時間のパターンなどが該当する。

3. 構造的側面からの証拠 得点の内的構造が構成概念の下位領域や次元性などの理論的構造に一致していることを示す証拠である。得点化手続きの適切性や項目間の相関関係のほか、因子分析の結果や内的整合性に関するデータも証拠に含まれる。

4. 一般化可能性の側面からの証拠 得点の意味や測定論的特性（平均や標準偏差、項目間の相関構造など）が、ある特定のデータセットだけでなく他の被験者集団、実施場面、実施時期、項目セットに対しても不変であるという証拠である。典型的には、一般化可能性理論による分散成分の検討結果が証拠となるが、再検査信頼性や平行検査信頼性、 α 係数も証拠となる。

5. 外的側面からの証拠 他の変数との間に、理論上想定される相関パターンが実際にも示されるという証拠である。同じ特性を測定していれば相関は高くなるはずであり（収束的証拠）、異なる特性を測定していれば相関は低くなるはずである（弁別的証拠）。また、同じ特性を測っていても、測定方法が異なればそれだけ相関は低めになるはずである。

6. 結果的側面からの証拠 妥当性は、特定の状況でその尺度を使用することの適切さを問題にする。したがって、その尺度を使用した結果生じた事態についても、妥当性評価の対象に組み込まれる。例えば、ある尺度を用いた結果、ある下位集団が系統的に不利になったとすれば、その場面でその尺度を使用したことが適切だったとは言い難い。短期的・長期的な悪影響が理論的・経験的に生じない、もしくは予見されないという証拠が必要である。

1.2 多特性多方法行列と収束的妥当性・弁別的妥当性

上述の6つの側面のうち、外的側面からの証拠として挙げられていた収束的証拠と弁別的証拠を体系的に検証するための有用な手法として、多特性多方法 (multitrait-multimethod; MTMM) 行列 (Campbell & Fiske, 1959) が用いられる。MTMM 行列とは、複数の構成概念 (“特性”) を複数の “方法” によって測定した際に得られる MTMM データから計算される相関行列のことである。

収束的妥当性 (convergent validity) は、同一の、あるいは類似した構成概念を測定している変数間で高い相関が得られることで示され、弁別的妥当性 (discriminant validity) は、異なる構成概念を測定している変数間で低い相関が得られることで示される (南風原, 2001)。Campbell & Fiske (1959) によると、弁別性の証拠は、構成概念の解釈で可能な対抗仮説を排除できるという意味で、特に重要である。MTMM 行列を利用することで、弁別的妥当性と収束的妥当性に関する検証を、強力かつ効果的に行うことができる (Messick, 1989)。

表 1.1 MTMM 行列の例 (Campbell & Fiske (1959) より引用)

	特性	方法 1			方法 2			方法 3			
		A	B	C	A	B	C	A	B	C	
方法 1	A	(.89)									
	B	.51	(.51)								
	C	.38	.37	(.76)							
方法 2	A	.57	.22	.09	(.93)						
	B	.22	.57	.10	.68	(.94)					
	C	.11	.11	.46	.59	.58	(.84)				
方法 3	A	.56	.22	.11	.67	.42	.33	(.94)			
	B	.23	.58	.12	.43	.66	.34	.67	(.92)		
	C	.11	.11	.45	.34	.32	.58	.58	.60	(.85)	

表 1.1 には、Campbell & Fiske (1959) から引用した MTMM 行列を掲載した。MTMM 行列の構成要素は、同一特性・同一方法 (monotrait-monomethod) の測定、同一特性・異方法 (monotrait-heteromethod) の測定、異特性・同一方法 (heterotrait-monomethod) の測定、異特性・異方法 (heterotrait-heteromethod) の測定という 4

つの部分に分けることができる。同一特性・同一方法の測定の部分は相関行列であれば1となるが、表 1.1 では丸括弧で囲みクロンバックの α 係数が記載されている。方法が異なっても同一の特性を測定している尺度間(同一特性・異方法)で相関が高ければ収束的妥当性の証拠となり、方法の異同にかかわらず異なる特性を測定している尺度間(異特性・同一方法および異特性・異方法)で相関が低ければ弁別的妥当性の証拠となると解釈される。

Campbell & Fiske (1959) は、収束的妥当性と弁別的妥当性について判断するための基準として、以下の4つを挙げている。

1. 同一特性・異方法の測定に対応するブロック行列の対角要素が有意に0とは異なり、十分に大きな値であれば、これにより収束的妥当性が保証される。
2. 同一特性・異方法の測定に対応する相関係数は、それに対応する各異方法ブロックの異特性・異方法三角行列の値よりも高くなくてはならない。同じ特性を測定している変数間の相関は、特性も方法も異にする他のどの変数との間の相関よりも高くなくてはいけない、ということである。
3. 同一特性・異方法の測定に対応する相関係数の値は、異特性・同一方法の相関よりも高くなくてはならない。これにより、たとえ同じ方法を用いても異なる特性を測定した場合よりも、方法にかかわらず同一特性を測定した場合の相関が高いということが示される。
4. 異特性相関について、どの方法の組み合わせにおいても特性ごとに同様のパターンが見られなくてはならない。

ただし、1は収束的妥当性の証拠であり、2~4は弁別的妥当性に関する証拠を与える。

1.3 多特性多方法行列のための確認的因子分析モデル

Campbell & Fiske (1959) が提案した多特性・多方法行列を用いた妥当性の検証は、手元の相関行列を観察し、目視によって上述の1~4までの条件を満たしてい

るか否かを確認するための指針である。その後、MTMM 行列を分析するための数々の統計モデルが提案された。特定の統計モデルを適用して MTMM 行列を分析することにより、妥当性についても目視に頼るのではなく、数値的な解釈を試みることができる。現在までに、分散分析 (analysis of variance, ANOVA; Boruch, Larkin, Wolins & MacKinney, 1970; Kavanagh, MacKinney & Wolins, 1971), 回帰分析 (regression analysis; Lehmann, 1988), パス解析 (path analysis; Kalleberg & Kluegel, 1975; Schmitt, 1978; Schmitt, Coyle, & Saari, 1977; Werts & Linn, 1970), 探索的因子分析 (exploratory factor analysis; Jackson, 1969; Jackson, 1975; Levin, 1988; Lomax & Algina, 1979) そして確認的因子分析 (Confirmatory factor analysis) など、MTMM 行列に対する様々な統計モデルの提案が行われてきた。

Schmitt, Coyle & Saari (1977) では、Campbell & Fiske (1959) が示した収束的妥当性と弁別的妥当性に関する条件を満たしているか否かを検証するためにそれまでに提案されてきた6種類の分析手法について比較検討している。結論として、個々の特性および方法を考慮して詳細な情報を提供するとともに、代替モデルについても評価できるという点でパス解析的手法 (Werts & Linn, 1970) が優れているとしている。また、分散分析モデル (Kavanagh et al., 1971) は、データの概要を的確に示してくれるという点では便利であり、相関行列内での比較には適しているが、各特性および方法が、Campbell & Fiske (1959) の指標と一致しているかを検討することはできないと指摘した。探索的因子分析に基づくモデル (Jackson, 1969 など) については、因子構造に興味があるのであればもっとも有用であると述べている。同様に、Schmitt & Stults (1986) もまた、MTMM データを対象とした分析方法について比較検討を行った。確認的因子分析モデルが主流となってきたとする一方で、直積モデル (direct product model; Browne, 1984; Bagozzi & Yi, 1990) の利点についても考察している。

直積モデルは、MTMM 行列の背後にある特性と方法は観測変数に対して乗法的に影響を与えているという Campbell & O'Connell (1967) の主張を、クロネッカー積を用いることで表現したモデルである。直積モデルは Cudeck (1988) や Wothke & Browne (1990) によって発展し、多くの下位モデルに分かれていった。Wothke & Browne (1990) は、それまで因子分析モデルとは別の枠組みで論じられてきた直積モデルを、2次因子分析モデルとして再表現した。Bagozzi & Yi

(1992)では直積モデルの下位モデルについて展望し、分析例を示した。Goffin & Jackson (1992)においてもいくつかの数値例が示された。直積モデルは、MTMMデータの分析で陥りやすいヘイウッドケースを回避するのに適しており、不適解が少ないことが経験的に知られている (Kenny & Kashy, 1992)。また、特性間、あるいは方法間の共分散構造、さらには誤差分散行列について柔軟に構造を入れることが可能である。ただし、特性と方法の影響が積の関係で表されているため、直積モデルでは特性因子と方法因子の説明割合を分離することができない。

現在では、MTMMデータの分析においては、特性と方法をそれぞれ因子として扱う確認的因子分析モデルが適用されることが多い (Marsh & Grayson, 1995; 豊田, 2000)。

モデル表現

本論文では、MTMM 行列を分析するための確認的因子分析モデルに焦点を絞って議論を進める。MTMM 行列に対して、確認的因子分析モデルを適用する最大の利点の1つは、観測変数の分散を特性、方法、誤差のそれぞれに分解することができるという点にある。一般的に、分解された分散成分は、観測変数に対する各要因による説明率を表している。そのため、MTMM 行列の文脈では、相対的に特性による説明率が高く、方法による説明率が低ければ、収束的妥当性と弁別的妥当性が高いと考えることができる。また、分散成分の分解により、信頼性係数についても古典的テスト理論の枠組みの延長で論じることができる。以上のように、確認的因子分析の母数の推定結果を利用して、信頼性や妥当性に関する指標を数値的に求めることが可能となる。

t 個の特性を m 個の方法によって測定した MTMM データに関する確認的因子分析モデルは以下のように表現される。

$$\mathbf{x} = \Lambda_T \mathbf{f}_T + \Lambda_M \mathbf{f}_M + \mathbf{e} \quad (1.1)$$

$\mathbf{x} (= (x_{11}, x_{21}, x_{31}, \dots, x_{ij}, \dots, x_{tm})')$ は $tm \times 1$ の観測変数ベクトルである。 x_{ij} によって、 i 番目の特性を j 番目の方法によって測定したときの観測変数を表す。

f_T は特性因子, f_M は方法因子をそれぞれ縦に並べたベクトル, e は誤差変数ベクトルである。

共分散構造分析 (structural equation modeling; SEM) の枠組みにおいて, 観測変数間の共分散構造行列 Σ は, 以下となる。

$$\Sigma = \Lambda_T \Phi_T \Lambda_T' + \Lambda_M \Phi_M \Lambda_M' + \Psi \quad (1.2)$$

ただし, Λ_T と Λ_M はそれぞれ特性因子と方法因子の因子負荷行列であり, 以下のように構成される。

$$\Lambda_T = \begin{bmatrix} \Lambda_{T_1} \\ \vdots \\ \Lambda_{T_j} \\ \vdots \\ \Lambda_{T_m} \end{bmatrix}, \Lambda_M = \begin{bmatrix} \lambda_{M_1} & 0 & 0 & \cdots & 0 \\ 0 & \lambda_{M_2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda_{M_m} \end{bmatrix} \quad (1.3)$$

Λ_{T_j} は $t \times t$ の対角行列 $\Lambda_{T_j} = \text{diag}(\lambda_{T_{ij}})$ で, $\lambda_{T_{ij}}$ は j 番目の方法で測定された i 番目の特性の観測変数が i 番目の特性因子 f_{T_i} から受ける影響の強さである。一方 λ_{M_j} は, i 番目の特性を j 番目の方法で測定した観測変数が j 番目の方法因子 f_{M_j} から受ける影響の強さ $\lambda_{M_{ij}}$ を要素としてもつ $t \times 1$ ベクトルである。 $t \times t$ 行列 Φ_T と $m \times m$ 行列 Φ_M は, それぞれ特性因子間, 方法因子間の相関行列であり, $E[ee'] = \Psi$ は対角要素に誤差分散 $\sigma_{e_{ij}}^2$ を配した $tm \times tm$ 対角行列である。

モデルの種類

以下では, 現在までに発展してきた MTMM 行列に対する確認的因子分析モデルのうち, 代表的な 4 種類のモデルについて, その概要と特徴を整理する。

CT-CM モデル

MTMM 行列のための, 最もシンプルで基礎となる確認的因子分析モデルが CT-CM (correlated trait-correlated method) モデル (Marsh, 1989; Marsh & Grayson, 1995; Widaman, 1985) である。CT-CM モデルは, 想定する特性と方法の数に対

応する t 個の特性因子と m 個の方法因子を仮定し、それらが独立に観測変数を説明するモデルである。ただし、特性因子と方法因子は互いに無相関である一方で、特性因子間、方法因子間にはそれぞれ相関を仮定する。すなわち、(1.2) 式の Φ_T と Φ_M の非対角要素を、いずれも自由母数として推定するモデルである。 $t = 3, m = 3$ の場合の CT-CM モデルのパス図を図 1.1 に示した。

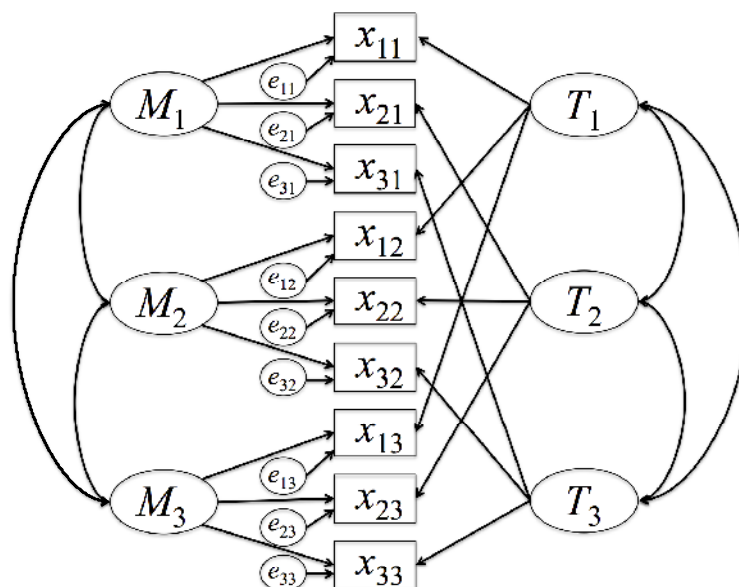


図 1.1 CT-CM モデル

このモデルは、特性と方法をそのまま因子として扱うため、モデルの仮定が明快である。また、1つのデータに対して必ず一通りのパラメタセットが得られるため、信頼性や妥当性の解釈が一通りに定まるというメリットがある。しかしながら、CT-CM モデルはしばしばモデルが識別不定になってしまうという欠点が報告されている。例えば、Wothke (1984) が 21 種類の MTMM 行列を分析したところ、そのうちの 10 種類の行列において識別不定に陥ったという。Grayson & Marsh (1994) では、通常の確認的因子分析モデルとして自由度に余裕のある CT-CM モデルでも、MTMM 行列の場合には生データの性質に依存して、因子負荷行列のランクが不足していると識別されないことが示されている。このような状況を踏まえ、Millsap (1992) では、MTMM データに対する確認的因子分析モデルにおいて回転が一意となるための十分条件を特定し、モデルが必ず識

別されるような制約の具体例を、適用例を用いて示している。また、確認的因子分析モデルによる MTMM 行列の分析では、分散が負の値として推定されるヘイウッドケースなど、不適解を生じやすいことも指摘されている (Marsh, 1989; Kenny & Kashy, 1992)。

CT-CU モデル

CT-CM モデルにおける不適解や識別の問題を回避するために有効とされるモデルの1つが、CT-CU (correlated trait-correlated uniqueness) モデル (Kenny, 1976; Lance, Noble & Scullen, 2002) である。このモデルは、図 1.2 のように、方法因子を仮定せず特性因子のみを扱い、同一方法による観測変数の誤差間と、特性因子間に相関を仮定する。(1.2) 式の共分散構造を $\Sigma = \Lambda_T \Phi_T \Lambda_T' + \Psi$ とし、 Ψ の非対角要素に適切に構造を仮定することで母数の推定を行うモデルである。

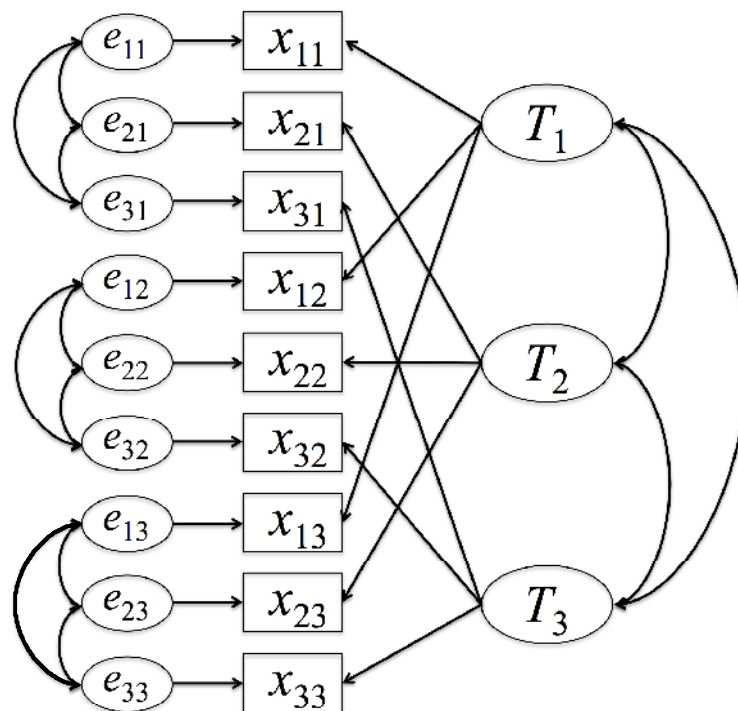


図 1.2 CT-CU モデル

CT-CU モデルでは、不適解と識別問題は改善されるものの、そもそも方法因子がモデルに組み込まれていないため、方法因子間の相関や方法因子と他の外

生変数との相関を仮定できないという短所がある。また、各方法に特有の効果
が、測定誤差（ランダム誤差）に含まれて交絡してしまうため、信頼性係数が
低く推定されるという報告もなされている。

CT-UM モデル

CT-UM (correlated trait-uncorrelated method) モデル (Grayson & Marsh, 1994;
Marsh & Grayson, 1995) は、CT-CM モデルに対して、方法因子間は無相関であ
るという制約を加えたモデルである。すなわち、(1.2) 式の Φ_M の対角要素を 1
とし、非対角要素はすべて 0 に固定する。パス図は図 1.3 の通りである。CT-UM
モデルは、Grayson & Marsh (1994) によって識別については証明されているが、
不適解の問題は完全に克服されてはいない (Marsh & Bailey, 1991)。

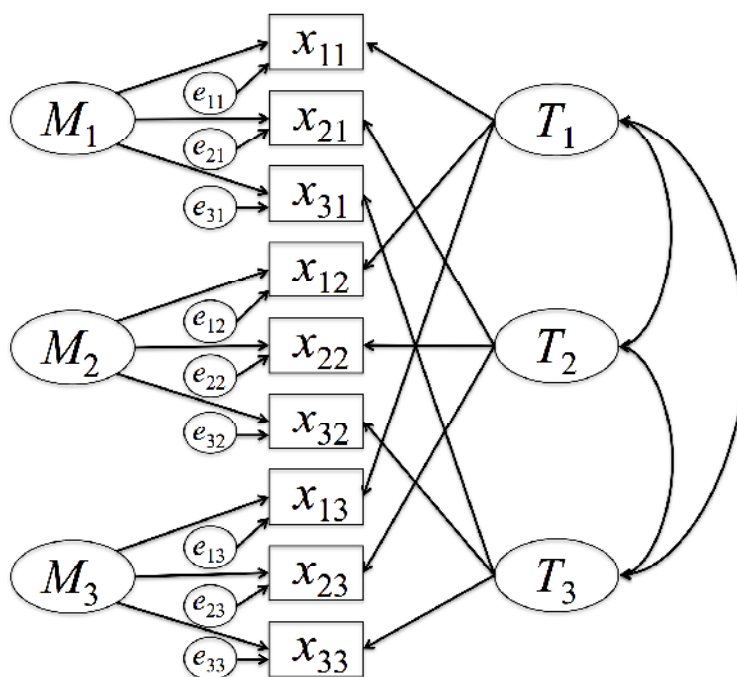


図 1.3 CT-UM モデル

CT-C(M-1) モデル

Eid (2000) は、識別不定の問題を克服するモデルとして CT-C(M-1) (correlated trait-correlated method minus one) モデル (Eid, Lischetzke, Nussbeck, & Triemer, 2003) を提案した。このモデルでは、 m 個の方法因子のうち、任意の 1 つの方法因子について基準として除外し、 $m - 1$ 個の方法因子を仮定するモデルである。例えば図 1.4 のように 1 つ目の方法を基準とすると、CT-C(M-1) モデルは、(1.3) 式中の λ_{M_1} の要素を全て 0 に固定し、さらに (1.2) 式の Φ_M のうち 1 行目および 1 列目の要素はすべて 0 に固定したモデルとして表現される。また、 $m - 1$ 個の方法因子間、および t 個の特性因子間にはそれぞれ相関を仮定する。

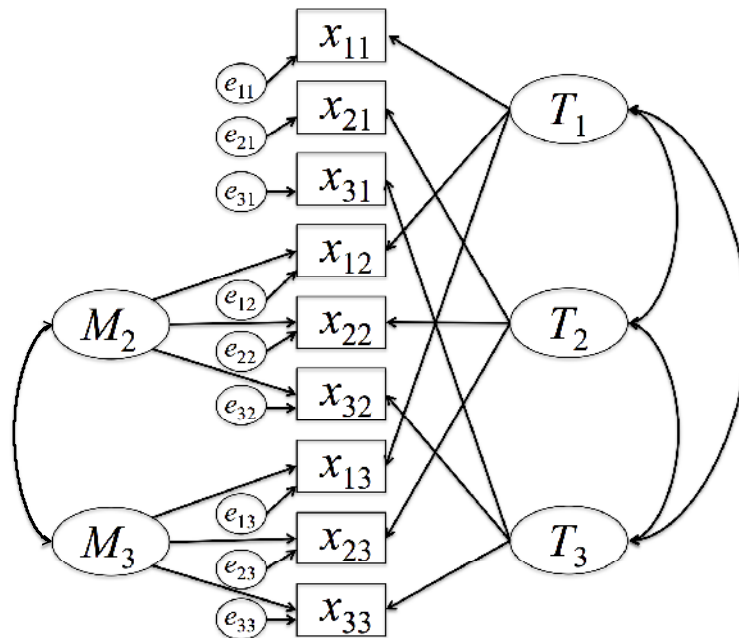


図 1.4 CT-C(M-1) モデル

Eid (2000) で、CT-C(M-1) モデルは、データとして少なくとも 3 つの特性と 3 つの方法が用いられていれば必ず識別されることが証明されている。しかしながら、この CT-C(M-1) モデルには、基準となる方法としてどの方法を指定するかによって得られるパラメタセットが変化し、これに伴い、同一データに対しても信頼性や妥当性に関する解釈が変わってしまうという欠点がある。

信頼性および収束的妥当性・弁別的妥当性に関する係数

Eid (2000) では、MTMM データに対する確認的因子分析モデルを古典的テスト理論の枠組みでとらえ、分散成分の分解を利用して、信頼性係数および収束的妥当性と弁別的妥当性を解釈するための係数を定義している。以下では、それらの係数について詳細を述べる。

古典的テスト理論では、観測変数は真値と誤差とに分解されると仮定する (Lord & Novick, 1968)。この仮定に基づき、(1.1) 式のモデルを個々の観測変数レベルで再定義すると、 i 番目の特性を j 番目の方法で測定した観測変数 x_{ij} は、真値 t_{ij} と誤差 e_{ij} の和として次のように表される。

$$x_{ij} = t_{ij} + e_{ij} \quad (1.4)$$

ただし、(1.4) 式の真値 t_{ij} には、特性に関する真値のみならず、方法因子による寄与部分も含まれる、すなわち、真値は次の (1.5) 式のように、特性因子の寄与部分 T_{ij} と方法因子の寄与部分 M_{ij} に分解される。

$$t_{ij} = T_{ij} + M_{ij} \quad (1.5)$$

$$= \lambda_{T_{ij}} f_{T_i} + \lambda_{M_{ij}} f_{M_j} \quad (1.6)$$

通常の古典的テスト理論では、タウ等価 (tau equivalent) 測定や平行 (parallel) 測定の条件を利用して測定精度の評価を行うため、データには一因子性を仮定している。しかしながら、(1.1) 式で表現される MTMM データのための確認的因子分析モデルでは、同じ特性因子から影響を受ける観測変数は、異なる方法因子からも影響を受けているため、一因子性の仮定が厳密には成立していない。そこで Eid (2000) は、(1.5) 式で表される真値の分解を、任意の 1 つの“方法”を比較の基準とすることで、次のように再定義した。ただしここでは、一般性を失わずに 1 番目の方法を基準とすることとする。

$$t_{ij} = E[t_{ij}|t_{i1}] + M_{ij} \quad (1.7)$$

(1.7) 式は、 x_{ij} に対応する真値 t_{ij} は、基準とした変数の真値 t_{i1} を用いて当該の真値 t_{ij} を回帰したときの予測値と、予測誤差 M_{ij} に分解されることを表して

いる。残差 M_{ij} は、真値 t_{ij} に含まれてはいるものの、当該変数と同じ特性を測定している観測変数の真値 t_{i1} からはまったく影響を受けない部分である。したがって M_{ij} は、1 番目の方法を基準としたときの方法 j 特有の効果と解釈される。また、先に述べた通り、 M_{ij} は t_{i1} を用いて t_{ij} を予測したときの残差なので、 t_{i1} と M_{ij} は独立である。その結果として、真値 t_{ij} の分散は

$$\text{var}(t_{ij}) = \text{var}(E[t_{ij}|t_{i1}]) + \text{var}(M_{ij}) \quad (1.8)$$

と分解することができる。

(1.8) 式で表される真値 t_{ij} の分散に占める、右辺第 1 項 $\text{var}(E[t_{ij}|t_{i1}])$ の割合

$$\text{Con}(t_{ij}) = \frac{\text{var}(E[t_{ij}|t_{i1}])}{\text{var}(t_{ij})}, \quad 0 < \text{var}(t_{ij}) < \infty \quad (1.9)$$

は、純粋に特性因子にのみ注目した場合の決定係数として解釈できる。(1.9) 式は、 x_{ij} の真値の分散に占める、当該変数と同じ特性を測定している観測変数の真値 t_{i1} でも説明できる部分の割合であり、方法によらない特性 i の一貫性を表している。したがって、“真値 t_{i1} に対する真値 t_{ij} の一貫性係数” (consistency coefficient of a true-score variable t_{ij} with respect to the true-score variable t_{i1}) と呼ばれる。Eid (2000) では、この係数が収束的妥当性を定量化した指標であると述べている。

これに対して、(1.8) 式の真値 t_{ij} の分散に占める、右辺第 2 項 $\text{var}(M_{ij})$ の割合

$$\text{MS}(t_{ij}) = \frac{\text{var}(M_{ij})}{\text{var}(t_{ij})}, \quad 0 < \text{var}(t_{ij}) < \infty \quad (1.10)$$

は、 x_{ij} の真値の分散のうち他の真値 t_{i1} では説明しきれない部分の割合である。 t_{ij} も t_{i1} も同じ特性 i について測定しているので、 $\text{MS}(t_{ij})$ は t_{ij} のうち特定の方法 j にのみ影響される部分の割合を意味しており、“真値 t_{i1} に対する真値 t_{ij} の方法特異性係数” (method specificity coefficient of a true-score variable t_{ij} with respect to the true score variable t_{i1}) と呼ばれる。

さらに(1.8) 式より、信頼性係数は以下のように定義される。

$$\text{Rel}(x_{ij}) = \frac{\text{var}(t_{ij})}{\text{var}(x_{ij})} = \frac{\text{var}(E[t_{ij}|t_{i1}])}{\text{var}(x_{ij})} + \frac{\text{var}(M_{ij})}{\text{var}(x_{ij})}, \quad 0 < \text{var}(x_{ij}) < \infty \quad (1.11)$$

(1.11)式の最右辺の第1項

$$\text{Con}(x_{ij}) = \frac{\text{var}(E[t_{ij}|t_{i1}])}{\text{var}(x_{ij})} \quad (1.12)$$

は“真値 t_{i1} に対する観測変数 x_{ij} の一貫性係数” (consistency coefficient of an observed variable x_{ij} with respect to the true-score variable t_{i1}) と呼ばれ、観測変数全体の分散に占める t_{ij} の条件付き平均の個人間分散を表す。

最後に、(1.11)式の最右辺の第2項

$$\text{MS}(x_{ij}) = \frac{\text{var}(M_{ij})}{\text{var}(x_{ij})} \quad (1.13)$$

は、観測変数全体の分散に占める方法特異な部分の割合であり、“真値 t_{i1} に対する観測変数 x_{ij} の方法特異性係数” (method specificity coefficient of an observed variable x_{ij} with respect to the true-score variable t_{i1}) と呼ばれる。

以上のように、古典的テスト理論の枠組みでMTMMデータのための確認的因子分析モデルを定義し、(1.9)式から(1.13)式までの係数を計算するためには、任意の1つの方法を比較の基準として選択する必要がある。上述の説明では、1つ目の方法を基準とした。本来、真値 t_{ij} は直接観測することのできない構成概念であり、その値を知ることはできない。そこでCT-C(M-1)モデルでは、比較の基準とする方法に対応する方法因子をモデルから除くことで、(1.14)式のように回帰モデルの観点から真値の予測値を得ることを可能にした。

$$E(t_{ij}|t_{i1}) = \mu_{ij} + \lambda_{T_{ij}}t_{i1} \quad (1.14)$$

ここで、比較基準として選ばれた方法に対応する真値 t_{i1} は、同一の添え字 i をもつすべての真値に影響を与える1つの共通因子と見なすことができる。CT-C(M-1)モデルでは、比較基準である1番目の方法によって測定された観測変数は、方法因子からの影響は受けずに、それぞれ1つの特性因子からのみパスを受けるため、 f_{T_i} の推定値をそのまま t_{i1} として解釈することが可能となる。

1.4 問題

モデルの汎用性

上述のように、MTMM 行列のための確認的因子分析モデルとして、CT-CM モデル、CT-CU モデル、CT-UM モデル、CT-C(M-1) モデルといった様々なモデルが提案されてきた。これらのモデルの優劣を決定する1つの方法として、Widaman (1985) では、段階的にネストしたモデルについて、 χ^2 値を比較することでモデルの適合を判断する方法を示している。比較する2つのモデルの構造がネストした関係である場合には、両モデルの適合度を示す χ^2 値の差も χ^2 分布に従い、かつ、このときの自由度は両モデルの自由度の差であるという性質を利用すると、モデル間の適合度の差を統計的に検定することができる (Kenny, 1976)。しかしながら、この方法は任意のデータとモデルとの当てはまりの良さを評価しているに過ぎない。汎用的にどのモデルが最も優れているか、という点については、それぞれに長所と短所があるため、議論の余地を残していると言えよう。

Eid (2000) によって CT-C(M-1) モデルが提案されるまで、MTMM 行列のための確認的因子分析モデルにおける議論の中心は専ら識別不定と不適解であった。このうち不適解は、モデルとデータとの適合が悪い場合に発生するものであり、当該母数の解釈はできなくなるものの、モデル全体の評価という観点においては有用な情報となり得る。これに対して、モデルが識別されるか否かは不適解の発生について議論する以前の問題であり、もし識別されなければ適合の吟味をはじめ、あらゆる解釈や考察が不可能となる。その意味で、識別不定に陥ることの方が重大な問題と言える。

CT-CM モデルの識別問題は、CT-CU モデルや CT-UM モデルによって解決される。しかし、これら2つのモデルでは、方法因子間の相関に対する研究仮説を反映することができない。CT-C(M-1) モデルは、識別問題と不適解の問題の両方に対して有効なモデルとして提案された。ところが、CT-C(M-1) モデルでは、どの方法を基準としてモデルから除くかによって、1つのデータセットに対して m 通りのパラメタセットが得られることとなる。同一のデータセットに対して複数通りの推定結果が得られるということは、測定の信頼性と妥当性に

関する解釈も一意に定まらないため、非常に不都合な性質である。以上のように、MTMM 行列に対する確認的因子分析モデルには、汎用的に利用できる決定版とも呼べるモデルがいまだに存在しない。

多特性多評価者データの分析

MTMM データにおける“方法”は、一般的に3つの種類に分けられる (Kenny, 1994)。1つ目は評価者 (rater)、2つ目は測定方法 (instrument-based)、3つ目は測定機会 (temporally-based) である。評価者を“方法”として扱う場合の MTMM データは特別に、多特性多評価者 (multitrait-multirater; MTMR) データ (Conway, 1996) と称される。MTMR データは主として、企業における人事評価場面で活用されている 360 度フィードバック (Edwards & Ewen, 1996a; Lepsinger & Lucia, 2009; 多面観察評価 (関本, 1965) ともいう) の結果として得られる。

360 度フィードバックは、1人の被評価者に対して上司や部下、同僚など周囲の人たちから仕事ぶりについてフィードバックしてもらい、本人の気付きを促し、人材育成や組織の活性化に役立てようというものである (相原・南雲, 2009)。日本では、人事考課の諸問題を克服するため 1965 年に関本昌秀が導入 (上田, 1974) して以来、360 度フィードバックはいくつかの企業で評価と育成を目的として独自に実施されてきた (馬場・馬場, 2005)。近年では、大企業を中心に導入する企業も増え、リクルートマネジメントソリューションズによる MOA、日本能率協会マネジメントセンターによる NAVI360 や RoundReview、日本経営協会総合研究所による多面観察<マネジメント版・リーダーシップ版>、そして産業能率大学総合研究所によるビジネス基礎力診断 S-BASE、プロ人材特性診断 S-ProII、マネジャー力診断 S-ATOM など、若手社員から中間管理職、経営者レベルまで、様々な立場を対象とした 360 度フィードバックのためのツールが数多く開発され、各社から提供されている。

このような現状を鑑みると、MTMR データにおける測定の信頼性および収束的妥当性、弁別的妥当性を検討することへの需要は決して低くはないだろう。さらに、人事アセスメントにおいては、測定の信頼性と妥当性が、アセスメントそのものの有効性を左右することになる (大沢, 2000)。したがって、360 度

フィードバックの結果として得られる MTMR 行列を用いて、測定の信頼性と妥当性について検討することは、人事アセスメントの質を吟味するという意味で非常に重要である。

MTMR データに対しても、先に述べたような確認的因子分析の枠組みにおける分析が行われてきた。例えば、Conway (1996) は、過去の文献から集めた 20 の MTMR 行列を分析対象として、CT-CM モデル、CT-CU モデル、直積モデルの 3 種類のモデルを適用し、識別と不適解、モデル適合の良さ、標準誤差の大きさといった観点から、どのモデルが望ましいかについて検討し、CT-CU モデルがもっとも優れていると結論付けている。また、Mount, Judge, Scullen, Sytsma & Hezlett (1998) では、2350 人の管理者を対象として、自己評価に加え、2 人の上司による評価、2 人の同僚による評価、2 人の部下による評価を行った MTMR データに確認的因子分析モデルを適用している。これらアメリカの先行研究の結果を受けて、日本における 360 度フィードバックのデータをもとに MTMR 行列を構成し、確認的因子分析モデルを適用して収束的妥当性および弁別的妥当性を検証した例として、高橋 (2001, 2010) が挙げられる。

以下では、MTMR データを分析対象として確認的因子分析モデルを適用する際に考慮すべき問題を 2 点指摘する。

合計得点における信頼性と妥当性の解釈

推定されたモデルの母数をそのまま解釈するだけにとどまらず、推定結果を利用して信頼性係数や妥当性係数を算出し、その値を解釈するという Eid (2000) のアイデアは画期的なものであった。MTMM 行列を用いた信頼性、および収束的妥当性・弁別的妥当性の検証においては、Campbell & Fiske (1959) に基づいた相関係数の解釈が礎となっているため、Eid (2000) による係数もまた、特性 i と方法 j との組み合わせで規定される 1 つ 1 つの観測変数 x_{ij} に対して計算される。

これに対して MTMR データの分析では、立場の異なる複数の評価者からの評定結果を特性ごとに合計し、各被評価者の得点（特性値）を計算したいという要求が考え得る。しかしながら、MTMM 行列に対する確認的因子分析の結果をもとに、特性ごとの合計得点の信頼性係数および妥当性に関する係数を導出し

た例はこれまでにはない。

同一立場内に複数の評価者がいる場合

MTMR データにおける、測定方法や測定機会を“方法”として扱う MTMM データとの大きな違いは、同一“方法”内で複数の測定値が得られる可能性があるという点である。例えば、1人の被評価者につき、複数の同僚、あるいは複数の部下から評定結果が得られる場合がある。また、得られる同僚評価の数、あるいは部下評価の数は、被評価者によって異なることが多い。

同一立場の評価者から複数の評定結果が得られた場合には、これまでその平均値を観測変数として用いて MTMM 行列が計算されてきた (Forsythe, McGaghie, & Friendman, 1986; Lawler, 1967; Mount, 1984)。平均値を用いてその他の変数との相関を求めると、真の値よりも希薄化することが知られている。そのため、同一立場の評価者が複数存在する場合に、単純に平均値を観測変数として用いてモデルを適用すると、分散成分の分解によって定義される信頼性係数や妥当性に関する係数は不正確になるだろう。また、評価者の数は測定の信頼性に直接的に寄与する要因の1つと考えられる。しかし、平均値を利用した分析では、被評価者ごとに異なる評価者の人数の情報を適切に反映した結果が得られない。MTMR データの分析においては、同一立場内で複数の評価者による評定が得られた場合の適切な分析方法について検討する必要がある。

1.5 研究目的および本稿の構成

本章では、妥当性研究の歴史と、MTMM 行列を用いて収束的妥当性および弁別的妥当性を検討するための確認的因子分析モデルについて概説した。さらに、代表的な4つのモデルを取り上げ、それぞれの特徴と問題点を整理した。本研究の目的は、MTMM データを用いた信頼性および妥当性の検討に利用される確認的因子分析モデルについて、その問題点を改善し、より実用的な分析方法を提案することである。また、単に確認的因子分析モデルを適用し、母数の推定を行うだけでなく、信頼性と妥当性を定量的に評価し、その解釈を行うま

での手順を手続き化して示す。本論文は、先の問題において述べた課題に対応する3つの研究から構成される。

まず第2章では、これまで提案されてきたモデルに比較して汎用性が高いと考えられる新たなモデルの提案を行う(研究I)。CT-CMモデルをベースとして、方法因子の因子得点の和を0とする制約を課すことで、どのようなデータに対しても適用可能で、識別の可能性も高く、信頼性と妥当性の解釈が一意に定まるモデルを開発した。

次に、第3章では、立場の異なる複数の評価者からの評定結果を特性ごとに合計した特性値について、その信頼性と収束的妥当性、弁別的妥当性を評価する方法について論じる(研究II)。ここでは、測定の信頼性や妥当性に影響を及ぼす要因として、テストや検査、質問紙調査の項目数に注目し、項目数の増減と信頼性係数および妥当性に関する係数の変化について一般式を導出した。この一般式を実データに対して適用することで、合計得点の信頼性と妥当性という観点から、評価者ごとに適切な項目配分を決定する方法について検討した。

第4章では、研究IIIで提案するマルコフ連鎖モンテカルロ(Markov chain Monte Carlo; MCMC)法によるアプローチについて理解を深めるため、ベイズ統計学およびMCMC法とその推定アルゴリズムについて概説した。

続く第5章では、同一立場内に複数の評価者が存在するMTMRデータを用いて、測定の信頼性と妥当性を適切に評価するための方法について検討した(研究III)。本研究においては、同一立場内に複数の評価者がいる場合に、その平均値を用いた分析において、評価者の人数に応じて分散成分がどのように異なるかを導出した。結果として、評価者の人数に応じて、誤差分散が異なることが明らかになった。そこで、評価者の人数に応じて被評価者ごとに異なる誤差分散の推定を可能にするために、MCMC法を利用する方法について提案した。

最後に、第6章では、これまでの研究を踏まえた研究全体の考察、および今後の課題について述べる。また付録として、研究において使用したデータおよび分析プログラムのスクリプトを掲載する。

第2章 研究I: 信頼性と妥当性の解釈が一意に定まるモデル

第2章では、MTMMデータに対する確認的因子分析モデルによるアプローチの問題点を踏まえ、それらを克服するために、新たなモデルを提案する。提案モデルは、特性因子と方法因子は無相関とし、方法因子の因子得点の和を0と置くモデルであり、信頼性と妥当性の解釈が必ず一意に定まるという点で、優れたモデルである。

本章では、はじめに、第1章で紹介したCT-CMモデルとCT-C(M-1)モデルの問題点を改めて整理した上で、提案モデルについて詳細な説明を行う。続いて、シミュレーション研究により、提案モデルの識別について検証を行った結果を示す。さらに、実データを用いた適用例により、提案モデルの適用可能性について検討する。適用例では、過去の文献から引用した12の相関行列を分析対象とし、それぞれのデータに対して提案モデル、CT-CMモデル、CT-C(M-1)モデルの3種類のモデルを適用した。推定結果について、モデルの識別と不適解の有無という2つの観点から比較するとともに、推定値を利用して、信頼性係数および収束的妥当性、弁別的妥当性に関する係数を計算し、各モデルの特徴について考察する。最後に、方法因子の因子得点の和を0とする制約を導入した提案モデルに関して、心理学研究において測定の信頼性および妥当性の検討に利用することの利点および、実用可能性という観点から論じる。

2.1 問題と目的

第1章でも述べた通り、MTMMデータに対する確認的因子分析モデルは主として、モデルの識別と不適解という2つの問題を抱えており、それらを克服することを目的として様々なモデルの提案がなされてきた。具体的には、CT-CMモ

デルを基本として、特性因子間および方法因子間の相関を仮定するか否か、方法因子をそもそも仮定するか否かといった観点から、CT-CUモデルやCT-UMモデルが考案された。もっとも基本的なモデルであるCT-CMモデルは、データの“特性”と“方法”をそのまま因子として表現しているという点で、自然な仮定によるわかりやすいモデル表現となっている。このモデルのメリットは、1つのデータに対して必ず一通りのパラメタセットが得られるため、信頼性や妥当性の解釈が一通りに定まるという点である。しかしながら、モデルの識別不定および不適解の発生が、実用における深刻な課題となっている。

これに対して、近年注目を集めているモデルがEid (2000)によるCT-C(M-1)モデルである。CT-C(M-1)モデルは、Eid (2000)の証明により以下の3つの条件下で識別が保証されている¹。ただし、 t と m はそれぞれデータにおける特性と方法の数、 Φ_T は $t \times t$ の特性因子間相関行列、 Φ_{M-1} は $(m-1) \times (m-1)$ の方法因子間相関行列である。

- (a) $t > 2$ かつ $m > 2$ 。
- (b) $t > 2, m = 2$ で、 Φ_T の要素がすべて自由母数で非ゼロの推定値となる。
- (c) $t = 2, m > 2$ で、 Φ_T と、 Φ_{M-1} の要素がすべて自由母数で非ゼロの推定値となる。

しかし、比較の基準とする方法を1つ選びその方法を除くことで、実際の方法の数より1つ少ない方法因子を仮定してモデルの推定を行うため、どの方法を基準として選択するかによって得られるパラメタセットが変化し、これに伴い、同一データに対しても信頼性や妥当性に関する解釈が変わってしまうという欠点がある。測定の信頼性や妥当性を検討するにあたって、同一のデータに対して結果が一意に定まらないことは望ましくない性質である。

一方で、因子負荷量を固定母数とし、等値制約を課するというアプローチによって識別に関する問題に対処したモデルとして、Kenny & Kashy (1992)による固定方法モデル (fixed method model) が挙げられる。固定方法モデルは、CT-C(M-1)

¹ここで挙げた3つの条件に加えて、一般的な確認的因子分析モデルの識別のための制約として、各因子から観測変数への因子負荷量のうち任意の1つを1に固定する、あるいは、因子の分散を1に固定するといういずれかの制約を課すものとする。

モデル開発のヒントとなったモデルでもあり、データにおける m 個の方法に対して $m-1$ 個の方法因子を仮定している。加えて、方法因子からの因子負荷量の合計が必ず 1 になるように母数を固定し、方法因子に関するサイズ $tm \times (m-1)$ の因子負荷行列 Λ_{M-1} について、以下のような制約を課す。

$$\Lambda_{M-1} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1} \\ -\mathbf{1} & -\mathbf{1} & -\mathbf{1} & \cdots & -\mathbf{1} \end{bmatrix} \quad (2.1)$$

ここで、 $\mathbf{1}$ はすべての要素が 1 である $t \times 1$ ベクトル、 $-\mathbf{1}$ はすべての要素が -1 である $t \times 1$ ベクトルを表す。すなわち、 $\lambda_{M_1}, \dots, \lambda_{M_{(m-2)}}$ までは、要素が 1 のみの 1 ベクトルとし、最後の $\lambda_{M_{(m-1)}}$ の要素は全て -1 に固定している。

固定方法モデルでは、方法因子の数を実際の方法の数より 1 つ減らし、方法因子からの因子負荷量の和が 0 になるように固定母数とすることで、モデルの自由度を増やし、識別の問題を改善している。しかしながら、Kenny & Kashy (1992) も結論で述べているように、方法因子の因子負荷量を 1 または -1 のいずれかしかとらない固定母数とすることは、あまりに強い制約である。実際には、同一の方法によって測定されている観測変数間で、方法因子からの因子負荷量がまったく同じ値になるとは限らない。さらに、固定方法モデルでは、方法の異なる観測変数間の因子負荷量にも等値の制約を置いていることになり、それぞれの方法が観測変数に与える影響の大きさの違いについては考慮しないモデルとなっている。

また、特性因子と方法因子の間にも相関を仮定しているため、特性と方法とに分散成分を分離して、信頼性と妥当性を検討することができない。心理学の領域におけるモデルの実用可能性という観点からも、特性因子と方法因子との相関は仮定しない方が適切であると考えられる。なぜなら、仮に特性因子と方法因子とに相関があるとすると、構成概念とそれを測定する尺度間に関係があるということの意味し、このような状況は想定し難いからである。

研究 I では、より自然な仮定の下に、信頼性および収束的妥当性・弁別的妥当性の解釈が 1 つのデータに対して一通りに定まるようなモデルを提案することを目的とする。提案モデルでは、因子負荷量ではなく、因子得点に制約を課

す。方法因子の因子得点の和が0になるという制約を導入することで、固定方法モデルやCT-C(M-1)モデルのように任意の1つの方法因子を削除しなくとも安定した推定が可能となり、データと推定結果が必ず一対一対応で定められる。また、特性因子と方法因子とは無相関とすることで、分散成分の分解を利用して収束的妥当性・弁別的妥当性の解釈を容易に行うことができる。

2.2 モデル

前章で説明した、 t 個の特性と m 個の方法を用いたMTMMデータのための確認的因子分析モデルの一般式、およびその共分散構造を再び以下に示す。

$$\mathbf{x} = \mathbf{\Lambda}_T \mathbf{f}_T + \mathbf{\Lambda}_M \mathbf{f}_M + \mathbf{e} \quad (2.2)$$

$$\mathbf{\Sigma} = \mathbf{\Lambda}_T \mathbf{\Phi}_T \mathbf{\Lambda}'_T + \mathbf{\Lambda}_M \mathbf{\Phi}_M \mathbf{\Lambda}'_M + \mathbf{\Psi} \quad (2.3)$$

ここで、方法因子の因子得点の和が0という制約を $m \times (m-1)$ の計画行列 \mathbf{k} を用いて表現すると、(2.2)式は(2.4)式に、(2.3)式は(2.5)式に書き換えられる。

$$\mathbf{x} = \mathbf{\Lambda}_T \mathbf{f}_T + \mathbf{\Lambda}_M \mathbf{k} \mathbf{f}_{M-1} + \mathbf{e} \quad (2.4)$$

$$\mathbf{\Sigma} = \mathbf{\Lambda}_T \mathbf{\Phi}_T \mathbf{\Lambda}'_T + \mathbf{\Lambda}_M \mathbf{k} \mathbf{\Phi}_{M-1} \mathbf{k}' \mathbf{\Lambda}'_M + \mathbf{\Psi} \quad (2.5)$$

計画行列 \mathbf{k} は、サイズ $m-1$ の単位行列の下にすべての要素が -1 の行ベクトルを結合した $m \times (m-1)$ の行列である。

$$\mathbf{k} = \begin{bmatrix} & \mathbf{I}_{(m-1)} & \\ -1 & \cdots & -1 \end{bmatrix} \quad (2.6)$$

\mathbf{f}_{M-1} は、計画行列 \mathbf{k} に対応して方法因子 \mathbf{f}_M の最後の1つを除いた因子得点ベクトルであり、同様に、 $\mathbf{\Phi}_{M-1}$ はサイズを $(m-1) \times (m-1)$ とした方法因子に関する分散共分散行列である。

例えば $t=3, m=3$ の場合、 $f_{M_1} + f_{M_2} + f_{M_3} = 0$ という制約は、以下のように表される。

$$\mathbf{f}_M = \mathbf{k} \mathbf{f}_{M-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} f_{M_1} \\ f_{M_2} \end{bmatrix} \quad (2.7)$$

ここで、(2.5)式の $\mathbf{k}\Phi_{M-1}\mathbf{k}'$ のみ書き下すと、

$$\begin{aligned} \mathbf{k}\Phi_{M-1}\mathbf{k}' &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} \phi_{M_1M_1} & \phi_{M_1M_2} \\ \phi_{M_1M_2} & \phi_{M_2M_2} \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} \phi_{M_1M_1} & \phi_{M_1M_2} & -\phi_{M_1M_1} - \phi_{M_1M_2} \\ \phi_{M_1M_2} & \phi_{M_2M_2} & -\phi_{M_2M_2} - \phi_{M_1M_2} \\ -\phi_{M_1M_1} - \phi_{M_1M_2} & -\phi_{M_2M_2} - \phi_{M_1M_2} & \phi_{M_1M_1} + \phi_{M_2M_2} + 2\phi_{M_1M_2} \end{bmatrix} \end{aligned} \quad (2.8)$$

となる。ただし、 $\phi_{M_1M_1} = \text{var}(f_{M_1})$, $\phi_{M_2M_2} = \text{var}(f_{M_2})$, $\phi_{M_1M_2} = \text{cov}(f_{M_1}, f_{M_2})$ である²。(2.5)式は $(m-1)$ 個の方法因子を仮定したモデルで $\mathbf{\Lambda}_M$ および Φ_{M-1} を推定することを意図しているが、(2.8)式より、それは m 個の方法因子を仮定したモデルにおいて因子間の分散と共分散に次のような制約を課して推定を行うことと同値であることがわかる。

$$\text{var}(f_{M_3}) = \text{var}(f_{M_1}) + \text{var}(f_{M_2}) + 2\text{cov}(f_{M_1}, f_{M_2}) \quad (2.9)$$

$$\text{cov}(f_{M_1}, f_{M_3}) = -\text{var}(f_{M_1}) - \text{cov}(f_{M_1}, f_{M_2}) \quad (2.10)$$

$$\text{cov}(f_{M_2}, f_{M_3}) = -\text{var}(f_{M_2}) - \text{cov}(f_{M_1}, f_{M_2}) \quad (2.11)$$

同様に、方法が4つの場合の計画行列 \mathbf{k} は

$$\mathbf{k} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix} \quad (2.12)$$

であり、因子間相関に対する制約は以下となる。

$$\begin{aligned} \text{var}(f_{M_4}) &= \text{var}(f_{M_1}) + \text{var}(f_{M_2}) + \text{var}(f_{M_3}) \\ &\quad + 2\text{cov}(f_{M_1}, f_{M_2}) + 2\text{cov}(f_{M_1}, f_{M_3}) + 2\text{cov}(f_{M_2}, f_{M_3}) \end{aligned} \quad (2.13)$$

$$\text{cov}(f_{M_1}, f_{M_4}) = -\text{var}(f_{M_1}) - \text{cov}(f_{M_1}, f_{M_2}) - \text{cov}(f_{M_1}, f_{M_3}) \quad (2.14)$$

$$\text{cov}(f_{M_2}, f_{M_4}) = -\text{var}(f_{M_2}) - \text{cov}(f_{M_1}, f_{M_2}) - \text{cov}(f_{M_2}, f_{M_3}) \quad (2.15)$$

$$\text{cov}(f_{M_3}, f_{M_4}) = -\text{var}(f_{M_3}) - \text{cov}(f_{M_2}, f_{M_3}) - \text{cov}(f_{M_1}, f_{M_3}) \quad (2.16)$$

²共分散を $\text{cov}()$ と表記した。

以上より、提案モデルでは、CT-C(M-1)モデルとは異なり、すべての方法因子について因子負荷量、そして分散と共分散が自由母数として推定される。また、(2.6)式では、一般性を失うことなく、最終行を-1のみのベクトルとしているが、実際に計画行列を作成する際には、任意の1行をすべて-1で構成される行ベクトルとし、それに対応するように f_{M-1} と Φ_{M-1} を用意すれば、解は一意に定まる。したがって、本モデルでは、信頼性と妥当性の解釈を必ず一通りに定めることができる。さらに、固定方法モデルのような因子負荷量に対する制約も必要としない。

因子得点に対する制約の意味

MTMM 行列を用いて信頼性および妥当性を検討するとき心理学者の興味の対象となるのは、方法ではなく、構成概念を表す特性の方である。方法因子の因子得点に対していかなる制約もない状況では、方法の違いによって特性の値にバラつきがあり、各被験者で方法の因子得点の合計は正になることも負になることもあり得る。ここで、因子得点の和が0という制約は、方法の因子得点の合計が正や負になるならば、それは当該被験者の特性として反映されるべきものと考え、測定を意図している特性に対して方法の影響を誤差と見なすという仮定を表現している。なお、本モデルの適用においては、方法間で尺度の単位が異なる場合を考慮して、標準化されたデータとして相関行列を分析対象とする。

信頼性と妥当性の解釈

本研究では、第1章で示した Eid (2000) の方法をもとに、分散成分の分解を利用してモデルの推定結果から信頼性と妥当性を解釈するためのより一般的な方法を提示する。まず、Eid (2000) と同様に、 i 番目の特性を j 番目の方法で測定した観測変数 x_{ij} は真値 t_{ij} と誤差 e_{ij} とに分解され、さらに真値 t_{ij} は特性因子の寄与部分と方法因子の寄与部分とに分けることができると仮定する。すな

わち,

$$x_{ij} = t_{ij} + e_{ij} \quad (2.17)$$

$$t_{ij} = \lambda_{T_{ij}} f_{T_i} + \lambda_{M_{ij}} f_{M_j} \quad (2.18)$$

である。ここで、誤差変数の期待値は0 ($E[e] = 0$)、かつ誤差変数と真値との積の期待値は0 ($E[te] = 0$) という古典的テスト理論の仮定および、そこから導き出される結果を利用すると、観測変数の分散は以下のように表される。

$$\text{var}(x_{ij}) = \text{var}(t_{ij}) + \text{var}(e_{ij}) \quad (2.19)$$

そして(2.18)式より、特性因子と方法因子を無相関とするモデルでは

$$\begin{aligned} \text{var}(t_{ij}) &= \lambda_{T_{ij}}^2 \text{var}(f_{T_i}) + \lambda_{M_{ij}}^2 \text{var}(f_{M_j}) \\ &= \text{var}(T_{ij}) + \text{var}(M_{ij}) \end{aligned} \quad (2.20)$$

となる。Eid (2000) では、観測不可能な真値 t_{ij} として、比較基準とした方法因子 f_{T_i} を用いるという方法により、信頼性係数および妥当性に関連する一貫性係数、方法特異性係数を定義していた。しかしながら、この方法はCT-C(M-1)モデルでしか利用することができないため、ここでは、モデルの母数として推定される(2.18)の右辺の値をそのまま真値 t_{ij} の予測値と見なし、信頼性係数や妥当性を解釈するための係数を計算する方法を示す。

最初に、(2.20)式から、観測変数の分散に占める真値の分散の割合を信頼性係数として定義する。

$$\text{Rel}(x_{ij}) = \frac{\text{var}(t_{ij})}{\text{var}(x_{ij})} = \frac{\text{var}(T_{ij})}{\text{var}(x_{ij})} + \frac{\text{var}(M_{ij})}{\text{var}(x_{ij})} \quad (2.21)$$

次に、(2.21)式の最終式の第1項、すなわち観測変数全体の分散に占める特性因子の変動による部分を観測変数の一貫性係数とする。

$$\text{Con}(x_{ij}) = \frac{\text{var}(T_{ij})}{\text{var}(x_{ij})} \quad (2.22)$$

これに対して(2.21)式の第2項

$$\text{MS}(x_{ij}) = \frac{\text{var}(M_{ij})}{\text{var}(x_{ij})} \quad (2.23)$$

は、観測変数全体の分散に占める方法因子の分散であり、観測変数の方法特異性係数とする。さらに、真値の分散に対する特性因子の分散 ((2.24) 式) は、異なる方法間での特性の一貫性として解釈できるため、真値の一貫性係数とする。

$$\text{Con}(t_{ij}) = \frac{\text{var}(T_{ij})}{\text{var}(t_{ij})} \quad (2.24)$$

測定方法が異なっても同一の特性で相関が高いことによって示される収束的妥当性は $\text{Con}(t_{ij})$ の値によって検証することができる。(2.18) 式のように真値は特性因子と方法因子の和であり、その真値の分散に占める特性因子の分散として定義される $\text{Con}(t_{ij})$ の値が大きいと、方法因子の異同にかかわらず特性因子の影響が大きいことを意味する。ただし、 $\text{Con}(x_{ij})$ が小さい場合には、そもそも観測変数全体の分散の中で特性因子の影響に依る部分が小さいということなので、 $\text{Con}(t_{ij})$ の値の大きさを解釈することはできない。

$\text{MS}(x_{ij})$ が高いということは、特性に関する差異が見出されずに、方法の違いに依存して観測変数の値が変動している状況である。したがって、 $\text{MS}(x_{ij})$ の値が小さいほど弁別的妥当性は高いと見なすことができる。また、信頼性は妥当性のための必要条件と考えられるため、 $\text{Rel}(x_{ij})$ が著しく低い場合には、 $\text{Con}(t_{ij})$ や $\text{MS}(x_{ij})$ の値にかかわらず、妥当性が低いと解釈する。

シミュレーションによる識別可能性の検証

提案モデルの識別の可能性について検証するため、シミュレーション研究を行った。特性の数および方法の数が増えると識別されやすくなるため、最も小さなサイズの MTMM 行列を想定し、 $t = 3, m = 3$ の相関行列を分析対象とした。

(2.5) 式で示した提案モデルについて、次のような等値制約を置き、表 2.1 のようにパラメタを設定した。ただし、方法因子間の共分散について $\phi_{M_{12}} = \phi_{M_{21}}$ である。

$$\begin{aligned}
\lambda_{T_{11}} &= \lambda_{T_{21}} = \lambda_{T_{31}} = \lambda_{T1} \\
\lambda_{T_{12}} &= \lambda_{T_{22}} = \lambda_{T_{32}} = \lambda_{T2} \\
\lambda_{T_{13}} &= \lambda_{T_{23}} = \lambda_{T_{33}} = \lambda_{T3} \\
\lambda_{M_{11}} &= \lambda_{M_{12}} = \lambda_{M_{13}} = \lambda_{M1} \\
\lambda_{M_{21}} &= \lambda_{M_{22}} = \lambda_{M_{23}} = \lambda_{M2} \\
\lambda_{T_{31}} &= \lambda_{M_{32}} = \lambda_{M_{33}} = \lambda_{M3} \\
\phi_{T_{12}} &= \phi_{T_{13}} = \phi_{T_{23}} = \phi_T \\
\phi_{T_{11}} &= \phi_{T_{22}} = \phi_{T_{33}} = \phi_{M_{11}} = 1
\end{aligned}$$

表 2.1 シミュレーションにおけるパラメタの設定

変数	Λ_T			Λ_M		
x_{11}	λ_{T1}	0	0	λ_{M1}	0	0
x_{21}	0	λ_{T1}	0	λ_{M2}	0	0
x_{31}	0	0	λ_{T1}	λ_{M3}	0	0
x_{12}	λ_{T2}	0	0	0	λ_{M1}	0
x_{22}	0	λ_{T2}	0	0	λ_{M2}	0
x_{32}	0	0	λ_{T2}	0	λ_{M3}	0
x_{13}	λ_{T3}	0	0	0	0	λ_{M1}
x_{23}	0	λ_{T3}	0	0	0	λ_{M2}
x_{33}	0	0	λ_{T3}	0	0	λ_{M3}
	Φ_T			Φ_M		
f_1	1	ϕ_T	ϕ_T	1	$\phi_{M_{12}}$	
f_2	ϕ_T	1	ϕ_T	$\phi_{M_{12}}$	$\phi_{M_{22}}$	
f_3	ϕ_T	ϕ_T	1			

表 2.1 に示したパラメタそれぞれについて、表 2.2 のような真値を与えた。すなわち、 $\lambda_{T1}, \lambda_{T2}, \lambda_{T3}$ はそれぞれ $\{0.3, 0.5, 0.7\}$ 、 $\lambda_{M1}, \lambda_{M2}, \lambda_{M3}$ はそれぞれ $\{0.2, 0.3, 0.4\}$ 、 ϕ_T は $\{0.3, 0.5, 0.7\}$ 、 $\phi_{M_{12}} (= \phi_{M_{21}})$ は $\{-0.7, -0.6, -0.5\}$ 、 $\phi_{M_{22}}$ は $\{0.8, 1.0, 1.2\}$ の値をとるものとし、合計 $3^9=19683$ パタンの相関行列を作成した。シミュレーションにより発生させた 19683 個の相関行列に対して、提案モ

表 2.2 各パラメタの真値

パラメタ	真値のパタン			パラメタ	真値のパタン			パラメタ	真値のパタン		
λ_{T1}	0.3	0.5	0.7	λ_{M1}	0.2	0.3	0.4	ϕ_T	0.3	0.5	0.7
λ_{T2}	0.3	0.5	0.7	λ_{M2}	0.2	0.3	0.4	ϕ_{M12}	-0.7	-0.6	-0.5
λ_{T3}	0.3	0.5	0.7	λ_{M3}	0.2	0.3	0.4	ϕ_{M22}	0.8	1.0	1.2

デルを適用して分析を行い，計算の収束について確認した。データ発生および分析には統計解析環境 R (Ihaka & Gentleman, 1996) 2.15.1 を用いた。母数の推定にはパッケージ lavaan を採用し，最適化アルゴリズムは BFGS 公式による準ニュートン法とした。

シミュレーション研究の結果，19683 個すべての相関行列において収束が確認された。したがって，実データで想定され得る MTMM 相関行列に対して，提案モデルは識別の可能性が非常に高いことが示唆された。

2.3 適用例

シミュレーション研究の結果から，提案モデルは識別の可能性が非常に高いことが示された。そこで適用例では，先行研究から引用した 12 の MTMM 行列 (実データ) を分析対象として，提案モデル，CT-CM モデル，CT-C(M-1) モデルをそれぞれ適用し，結果について比較することで，提案モデルの実用性についてさらなる検討を行う。

方法

提案モデルを実際の相関行列に適用し，信頼性係数および妥当性に関連する一貫性係数や方法特異性係数を算出する。本章の冒頭で述べた通り，データにおける特性と方法がそれぞれ 3 つ以上であれば，CT-C(M-1) モデルは確実に識別可能であることが Eid (2000) で証明されている。そこで，考察においてモデル間の比較を行うことを考慮して，特性と方法の数がともに 3 つ以上であるという条件を満たす 12 の MTMM 行列を先行研究より引用し，分析対象として用

いた³。

分析には EQS6.1 for Windows ならびに Mplus (ver.4.1) を使用した。EQS ではモデルが識別可能ならば、ハイウッドケースの場合にも推定を途中で停止することなく負に推定された分散の推定値を 0 に固定してモデルの推定が行われる。よって、EQS による推定が成功した場合には識別されたと見なすことができる。さらに Mplus を用いることで、ハイウッドケースが実際に生じていることを確認し、不適解の判断を行った。

結果と考察

実データの分析において、CT-CM モデル、CT-C(M-1) モデル、方法因子の因子得点の和が 0 という制約を課した提案モデルの 3 種類を、12 の MTMM 行列に適用した結果の解の状態を表 2.3 に示した。

表 2.3 3 種類の各モデルにおける推定結果の解の状態

データ	CT-CM モデル	CT-C(M-1) モデル	提案モデル
Biesanz & West (2004)	正常	正常	正常
Carroll (1952)	識別不定	不適解	正常
Coenders & Saris (2000)	正常	正常	正常
Kelly & Fiske (1951)	識別不定	不適解	不適解
Kothandapani (1971)	不適解	不適解	不適解
Levin (1973)	正常	正常	正常
Maas, Lensvelt-Mulders & Hox (2009)	正常	正常	正常
Manson, Levine & Brannick (2000)	不適解	不適解	不適解
Muis, Winne & Jamieson-Noel (2007)	識別不定	不適解	不適解
Ostrom (1969)	不適解	不適解	不適解
Roshal, Frieze & Wood (1971)	不適解	不適解	不適解
Steger, Frazier, Oishi & Kaler (2006)	不適解	正常	正常

CT-C(M-1) モデルではどの方法を基準とするかによって解の状態が異なることがあるが、 m 通りすべての場合で推定を行い、その結果どの方法を除いた場

³分析に使用した MTMM 行列は、以下の 12 の論文からの引用である。Biesanz & West (2004), Carroll (1952), Coenders & Saris (2000), Kelly & Fiske (1951), Kothandapani (1971), Levin (1973), Maas, Lensvelt-Mulders & Hox (2009), Manson, Levine & Brannick (2000), Muis, Winne & Jamieson-Noel (2007), Ostrom (1969), Roshal, Frieze & Wood (1971), Steger, Frazier, Oishi & Kaler (2006)。

合でも必ず識別され、かつ正常な解が得られた場合のみ“正常”として扱った⁴。また、不適解には、因子の分散あるいは誤差分散が負になる、因子負荷量が桁外れに大きな値になる場合を含む。

表 2.3 から、不適解はモデルの種類に関わらず、だいたい同じデータで生じていることがわかる。現実には得られるデータは必ずしも仮定した共分散構造モデルに従って生成されるとは限らないため、モデルとデータが合致していない場合に不適解が生じる。すなわち、不適解が生じた相関行列は、データの発生過程がモデルで仮定している共分散構造とは異なるものであったという可能性が示唆されたことになる。一方で、識別不定はモデルに原因があって生じる問題であり、モデルが識別されていない場合には解を得ることができないため、より根本的な問題である。CT-CM モデルでは、今回使用した 12 の MTMM 行列のうち 3 つの MTMM 行列においてモデルが識別されなかった。これに対して、方法因子の因子得点の和が 0 という制約を課した提案モデルでは、今回分析を行った 12 の相関行列において識別されない例はなかった。

そこで、3 種類のモデルすべてにおいて正常な解が得られた 4 つのデータについて、(2.21) 式から (2.24) 式までを用いて、 $Rel(x_{ij})$, $Con(x_{ij})$, $MS(x_{ij})$, $Con(t_{ij})$ を算出した結果をそれぞれ表 2.4 から表 2.7 に示した⁵。表の左半分には、CT-C(M-1) モデルにおいて基準とする方法を順番に変えて m 通りの係数を求めた結果を示し、右半分には、CT-CM モデルと提案モデルの結果を示した。CT-C(M-1) モデルでは、相対的な大小関係は保たれているものの、どの方法を基準として選ぶかによってすべての係数において異なる値が算出された。

ビッグファイブの 5 つの特性について、自己評価による測定を経時的に 3 回繰り返した Biesanz & West (2004) の MTMM 行列 (表 2.4) では、いずれのモデ

⁴同一の MTMM 行列に対して同じモデルを適用しているにもかかわらず、基準とする方法の選び方によって正常な解が得られる場合と不適解となる場合があることが問題であると考えたため。また、不適解についても同様に、どの方法を基準とするかによって不適解が生じる場合と生じない場合がある。基準として除く方法を順に変えて m 通りすべての場合で推定を行い、そのすべての場合を通して 1 つでも不適解があったものを表 2.3 では不適解として報告している。

⁵本研究では、CT-C(M-1) モデルに対しても、(2.21) 式から (2.24) 式を用いて係数を算出した。なお、 $Rel(x_{ij})$, $Con(x_{ij})$, $MS(x_{ij})$, $Con(t_{ij})$ を求める際には、これ以後すべて EQS による推定結果を利用している。また、表の直下には各 MTMM 行列で用いられている特性および方法の内容を示した。表中の x_{ij} は、 i 番目の特性を j 番目の方法により測定した観測変数を表している。したがって、例えば表 2.4 において、 x_{23} は、“誠実さ”という特性の“自己評価 3 回目”の結果に対応する観測変数である。

ルでも信頼性係数はすべて0.7以上と一貫して比較的高い値であった。方法因子の影響としての $MS(x_{ij})$ が0.1を超える変数は少なく、 $Con(x_{ij})$ との差が大きいことから特性の変動が観測変数に与える影響が大きいことがわかる。表2.5に示したCoenders & Saris (2000)の結果では、相対的に M_1 による測定の信頼性が低い。妥当性については、弁別的妥当性、収束的妥当性ともに高いと解釈できるが、提案モデルとCT-CMモデルでは、ほとんどの変数で $Con(x_{ij})$ の値よりも $MS(x_{ij})$ の値の方が1桁小さく、 $Con(t_{ij})$ も総じて0.8以上であったのに対して、CT-C(M-1)モデルの結果では $MS(x_{ij})$ の値が大きく、その影響で $Con(t_{ij})$ の値が小さくなっていることがわかる。表2.6では、モデル間での係数の値の違いが小さく、信頼性係数の値はすべて0.8以上である。さらに、一貫して $Con(x_{ij})$ と $Con(t_{ij})$ の値が大きく $MS(x_{ij})$ の値が小さい。よって、Maas et al. (2009)の調査は測定の精度が非常に高いと言える。

表2.4から表2.6までの3つのMTMM行列では、信頼性係数の値が総じて高く、必ず $Con(x_{ij}) > MS(x_{ij})$ が成り立っている。これに対して表2.7に結果を示したLevin (1973)のMTMM行列では、一見CT-C(M-1)モデルで基準とする変数を変えても信頼性係数の値に大きな違いはないように見えるが、 $Con(t_{ij})$ の値の変化が大きい。すなわち、どの方法を除いたかによって $Con(x_{ij})$ と $MS(x_{ij})$ の割合に変化が見られ、例えば x_{12} や x_{22} では、 M_1 を基準としたときには $Con(x_{ij}) < MS(x_{ij})$ であるが M_3 を基準としたときには $Con(x_{ij}) > MS(x_{ij})$ となっており、このような大小関係の逆転は他の観測変数においても生じている。特に、オブザーバーからの評価(M_3)を除いた場合に、 $MS(x_{ij})$ の値は小さく、かつ $Con(x_{ij})$ の値が大きくなる傾向が見られた。Levin (1973)では、 x_{31} や x_{33} で $Rel(x_{ij})$ の値が0.5を下回っており、特に T_3 (帰属度)の測定、および M_1 (自己評価)において信頼性が著しく低いことがわかる。また、表2.7の右側より、Levin (1973)のMTMM行列では、提案モデルとCT-CMモデルとの間でも、 $Con(x_{ij})$ 、 $MS(x_{ij})$ 、 $Con(t_{ij})$ の値に大きな違いが観察された。

表 2.4 3 種類のモデルによる結果の比較 (Biesanz & West, 2004)

変数	Rel(x_{ij})	Con(x_{ij})	MS(x_{ij})	Con(t_{ij})	Rel(x_{ij})	Con(x_{ij})	MS(x_{ij})	Con(t_{ij})
CT-C(M-1) モデル ($M1$ 基準)					CT-CM モデル			
x_{11}	0.834				0.821	0.666	0.155	0.811
x_{21}	0.822				0.857	0.682	0.175	0.796
x_{31}	0.787				0.790	0.766	0.024	0.970
x_{41}	0.723				0.727	0.665	0.062	0.915
x_{51}	0.806				0.821	0.735	0.086	0.895
x_{12}	0.935	0.769	0.167	0.822	0.943	0.617	0.326	0.654
x_{22}	0.867	0.780	0.086	0.900	0.882	0.745	0.137	0.845
x_{32}	0.889	0.807	0.082	0.908	0.888	0.796	0.092	0.896
x_{42}	0.970	0.650	0.221	0.747	0.868	0.576	0.292	0.664
x_{52}	0.882	0.799	0.082	0.906	0.884	0.816	0.069	0.923
x_{13}	0.792	0.529	0.263	0.668	0.778	0.531	0.247	0.683
x_{23}	0.794	0.602	0.193	0.758	0.782	0.622	0.160	0.795
x_{33}	0.826	0.707	0.120	0.856	0.833	0.695	0.138	0.834
x_{43}	0.796	0.642	0.155	0.807	0.798	0.607	0.191	0.761
x_{53}	0.849	0.742	0.107	0.874	0.844	0.750	0.094	0.889
CT-C(M-1) モデル ($M2$ 基準)					提案モデル			
x_{11}	0.832	0.595	0.237	0.715	0.854	0.733	0.121	0.858
x_{21}	0.832	0.684	0.148	0.822	0.838	0.768	0.071	0.916
x_{31}	0.784	0.719	0.065	0.917	0.779	0.766	0.013	0.983
x_{41}	0.736	0.610	0.126	0.829	0.708	0.669	0.040	0.945
x_{51}	0.820	0.737	0.082	0.899	0.813	0.787	0.026	0.968
x_{12}	0.954				0.943	0.844	0.100	0.895
x_{22}	0.889				0.870	0.839	0.030	0.964
x_{32}	0.895				0.896	0.861	0.034	0.961
x_{42}	0.845				0.869	0.756	0.113	0.870
x_{52}	0.891				0.877	0.852	0.024	0.971
x_{13}	0.773	0.593	0.181	0.767	0.765	0.675	0.089	0.882
x_{23}	0.782	0.655	0.127	0.838	0.795	0.706	0.089	0.888
x_{33}	0.832	0.731	0.101	0.879	0.825	0.768	0.057	0.931
x_{43}	0.822	0.627	0.195	0.763	0.793	0.717	0.076	0.904
x_{53}	0.846	0.769	0.077	0.909	0.849	0.804	0.045	0.947
CT-C(M-1) モデル ($M3$ 基準)								
x_{11}	0.838	0.584	0.254	0.697				
x_{21}	0.845	0.673	0.172	0.796				
x_{31}	0.787	0.723	0.063	0.919				
x_{41}	0.740	0.566	0.174	0.765				
x_{51}	0.820	0.739	0.081	0.901				
x_{12}	0.945	0.691	0.254	0.731				
x_{22}	0.882	0.767	0.116	0.870				
x_{32}	0.896	0.799	0.097	0.892				
x_{42}	0.859	0.668	0.191	0.778				
x_{52}	0.886	0.806	0.080	0.910				
x_{13}	0.783							
x_{23}	0.776							
x_{33}	0.805							
x_{43}	0.800							
x_{53}	0.843							

T_1 : 調和性 T_2 : 誠実さ T_3 : 外向性 T_4 : 神経症傾向 T_5 : 開放性
 M_1 : 自己評価 1 回目 M_2 : 自己評価 2 回目 M_3 : 自己評価 3 回目

表 2.5 3種類のモデルによる結果の比較 (Coenders & Saris, 2000)

変数	Rel(x_{ij})	Con(x_{ij})	MS(x_{ij})	Con(t_{ij})	Rel(x_{ij})	Con(x_{ij})	MS(x_{ij})	Con(t_{ij})
	CT-C(M-1)モデル (M1 基準)				CT-CMモデル			
x_{11}	0.677				0.712	0.584	0.128	0.820
x_{21}	0.798				0.834	0.731	0.102	0.876
x_{31}	0.726				0.650	0.603	0.047	0.928
x_{12}	0.692	0.550	0.142	0.795	0.737	0.731	0.005	0.992
x_{22}	0.800	0.757	0.043	0.946	0.763	0.763	0.000	1.000
x_{32}	0.853	0.566	0.287	0.664	0.822	0.816	0.006	0.993
x_{13}	0.783	0.625	0.158	0.798	0.745	0.664	0.081	0.891
x_{23}	0.875	0.797	0.078	0.911	0.884	0.804	0.079	0.910
x_{33}	0.874	0.625	0.249	0.715	0.900	0.737	0.163	0.819
	CT-C(M-1)モデル (M2 基準)				提案モデル			
x_{11}	0.731	0.478	0.253	0.654	0.692	0.615	0.076	0.889
x_{21}	0.819	0.687	0.132	0.839	0.847	0.759	0.088	0.896
x_{31}	0.689	0.573	0.166	0.832	0.652	0.623	0.029	0.956
x_{12}	0.719				0.755	0.654	0.100	0.866
x_{22}	0.812				0.786	0.781	0.005	0.994
x_{32}	0.794				0.804	0.728	0.075	0.905
x_{13}	0.775	0.592	0.182	0.764	0.742	0.689	0.052	0.929
x_{23}	0.865	0.755	0.110	0.873	0.875	0.820	0.055	0.937
x_{33}	0.879	0.634	0.245	0.721	0.913	0.759	0.154	0.831
	CT-C(M-1)モデル (M3 基準)							
x_{11}	0.663	0.535	0.129	0.807				
x_{21}	0.820	0.724	0.096	0.883				
x_{31}	0.743	0.548	0.195	0.738				
x_{12}	0.743	0.521	0.222	0.701				
x_{22}	0.806	0.739	0.067	0.917				
x_{32}	0.774	0.596	0.179	0.770				
x_{13}	0.815							
x_{23}	0.869							
x_{33}	0.860							

T_1 : 生活全般の満足度 T_2 : 経済状況の満足度 T_3 : 社会的接触の満足度性
 M_1 : 100点満点評価 M_2 : 5段階評価 M_3 : 11段階評価

表 2.6 3 種類のモデルによる結果の比較 (Maas et al., 2009)

変数	Rel(x_{ij})	Con(x_{ij})	MS(x_{ij})	Con(t_{ij})	Rel(x_{ij})	Con(x_{ij})	MS(x_{ij})	Con(t_{ij})
	CT-C(M-1)モデル (M1 基準)				CT-CMモデル			
x_{11}	0.892				0.901	0.766	0.135	0.850
x_{21}	0.923				0.947	0.695	0.252	0.734
x_{31}	0.923				0.924	0.843	0.081	0.912
x_{12}	0.903	0.803	0.099	0.889	0.901	0.739	0.163	0.820
x_{22}	0.925	0.823	0.102	0.890	0.949	0.872	0.078	0.918
x_{32}	0.938	0.879	0.059	0.937	0.935	0.879	0.055	0.941
x_{13}	0.899	0.800	0.100	0.890	0.914	0.689	0.225	0.753
x_{23}	0.918	0.768	0.150	0.837	0.880	0.692	0.187	0.786
x_{33}	0.838	0.786	0.052	0.938	0.836	0.774	0.062	0.992
	CT-C(M-1)モデル (M2 基準)				提案モデル			
x_{11}	0.884	0.787	0.096	0.890	0.886	0.843	0.042	0.951
x_{21}	0.933	0.796	0.137	0.853	0.932	0.871	0.062	0.935
x_{31}	0.928	0.866	0.062	0.933	0.925	0.900	0.025	0.973
x_{12}	0.894				0.903	0.870	0.033	0.863
x_{22}	0.935				0.916	0.896	0.021	0.978
x_{32}	0.933				0.939	0.921	0.018	0.981
x_{13}	0.921	0.811	0.109	0.881	0.890	0.860	0.031	0.966
x_{23}	0.885	0.799	0.086	0.903	0.917	0.857	0.060	0.935
x_{33}	0.834	0.802	0.031	0.962	0.832	0.819	0.013	0.984
	CT-C(M-1)モデル (M3 基準)							
x_{11}	0.896	0.776	0.120	0.866				
x_{21}	0.928	0.797	0.131	0.859				
x_{31}	0.928	0.863	0.065	0.930				
x_{12}	0.891	0.823	0.068	0.924				
x_{22}	0.933	0.846	0.087	0.907				
x_{32}	0.936	0.892	0.045	0.953				
x_{13}	0.894							
x_{23}	0.893							
x_{33}	0.839							

T_1 : 住まいの満足度 T_2 : 収入の満足度 T_3 : 健康の満足度
 M_1 : はしごグラフ M_2 : 顔型グラフ M_3 : 円グラフ

表 2.7 3種類のモデルによる結果の比較 (Levin, 1973)

変数	Rel(x_{ij})	Con(x_{ij})	MS(x_{ij})	Con(t_{ij})	Rel(x_{ij})	Con(x_{ij})	MS(x_{ij})	Con(t_{ij})
	CT-C(M-1) モデル (M1 基準)				CT-CM モデル			
x_{11}	0.598				0.596	0.168	0.428	0.282
x_{21}	0.538				0.539	0.078	0.460	0.145
x_{31}	0.384				0.375	0.011	0.363	0.029
x_{41}	0.755				0.744	0.236	0.508	0.317
x_{12}	0.805	0.392	0.413	0.487	0.839	0.260	0.579	0.310
x_{22}	0.756	0.268	0.488	0.354	0.786	0.016	0.770	0.020
x_{32}	0.615	0.169	0.446	0.275	0.686	0.009	0.678	0.013
x_{42}	0.923	0.355	0.568	0.385	0.903	0.145	0.758	0.161
x_{13}	0.907	0.453	0.454	0.499	0.911	0.295	0.616	0.324
x_{23}	0.845	0.389	0.456	0.460	0.912	0.269	0.643	0.295
x_{33}	0.474	0.236	0.238	0.498	0.484	0.057	0.427	0.118
x_{43}	0.969	0.446	0.523	0.460	0.975	0.254	0.721	0.261
	CT-C(M-1) モデル (M2 基準)				提案モデル			
x_{11}	0.596	0.228	0.367	0.383	0.583	0.441	0.142	0.756
x_{21}	0.526	0.145	0.380	0.276	0.543	0.346	0.196	0.637
x_{31}	0.320	0.116	0.203	0.363	0.339	0.231	0.108	0.681
x_{41}	0.714	0.292	0.422	0.409	0.714	0.537	0.176	0.752
x_{12}	0.827				0.813	0.693	0.120	0.852
x_{22}	0.773				0.772	0.583	0.189	0.755
x_{32}	0.636				0.652	0.423	0.229	0.649
x_{42}	0.925				0.915	0.746	0.169	0.815
x_{13}	0.906	0.569	0.337	0.628	0.914	0.809	0.105	0.885
x_{23}	0.846	0.410	0.437	0.485	0.855	0.703	0.152	0.822
x_{33}	0.466	0.357	0.109	0.766	0.470	0.450	0.019	0.957
x_{43}	0.965	0.545	0.420	0.565	0.967	0.826	0.141	0.854
	CT-C(M-1) モデル (M3 基準)							
x_{11}	0.598	0.308	0.290	0.515				
x_{21}	0.528	0.225	0.304	0.426				
x_{31}	0.332	0.146	0.186	0.440				
x_{41}	0.722	0.365	0.358	0.506				
x_{12}	0.801	0.533	0.268	0.665				
x_{22}	0.766	0.409	0.356	0.534				
x_{32}	0.628	0.289	0.339	0.460				
x_{42}	0.912	0.565	0.347	0.620				
x_{13}	0.935							
x_{23}	0.863							
x_{33}	0.490							
x_{43}	0.964							

T_1 : 目立ち度 T_2 : 貢献度 T_3 : 帰属度 T_4 : リーダー度

M_1 : 自己評価 M_2 : 仲間からの評価 M_3 : オブザーバーからの評価

表 2.8 と表 2.9 には、先の 4 つのデータの他に提案モデルにおいて正常な解が得られた Carroll (1952) と、Steger et al. (2006) に対する結果を示した。

表 2.8 提案モデルの結果 (Carroll, 1952)

	Rel(x_{ij})	Con(x_{ij})	MS(x_{ij})	Con(t_{ij})
x_{11}	0.913	0.842	0.072	0.922
x_{21}	0.724	0.627	0.097	0.866
x_{31}	0.980	0.653	0.327	0.666
x_{41}	0.939	0.635	0.304	0.676
x_{51}	0.865	0.863	0.001	0.998
x_{12}	0.476	0.405	0.071	0.851
x_{22}	0.728	0.653	0.075	0.897
x_{32}	0.488	0.444	0.044	0.910
x_{42}	0.426	0.415	0.011	0.974
x_{52}	0.460	0.113	0.347	0.246
x_{13}	0.434	0.248	0.186	0.571
x_{23}	0.504	0.308	0.196	0.611
x_{33}	0.435	0.265	0.170	0.609
x_{43}	0.163	0.146	0.017	0.896
x_{53}	0.602	0.194	0.408	0.322

T_1 : 社会的内向性 T_2 : 思考的内向性 T_3 : 抑うつ T_4 : 循環気質傾向 T_5 : のんきさ
 M_1 : ギルフォード検査 M_2 : 自己評価 M_3 : クラスメイトからの評価

表 2.8 の Carroll (1952) の結果では、自己評価やクラスメイトからの評価よりもギルフォード検査の信頼性が高かった。また、4 番目の特性 (循環気質傾向) に注目して Con(x_{ij}) と Con(t_{ij}) とを比較すると、 x_{41} は Con(t_{ij}) の値が最も小さいものの、Con(x_{ij}) の値は 0.635 と最も大きく、一方で Con(t_{ij}) の値が大きい x_{42} 、 x_{43} ではそれぞれ Con(x_{42}) = 0.415、Con(x_{43}) = 0.146 と小さな値となっている。 x_{42} と x_{43} は Rel(x_{ij}) も非常に小さく、そもそも測定の信頼性が著しく低いことがわかる。Con(t_{ij}) の値が高くとも、Con(x_{ij}) の値が小さければ、誤差の分散があまりに大きく、特性因子や方法因子の分散成分の割合が小さいため、特性因子や方法因子による影響はもとより微々たるものであることを意味する。Con(x_{ij}) の値はどの特性に対してもだいたい $M_1 > M_2 > M_3$ となっており、Con(t_{ij}) の値だけを見て自己評価は収束的妥当性が高いと判断するのは誤りである。特に、自己評価 (M_2) とクラスメイトからの評価 (M_3) では

信頼性係数がほとんど 0.5 を下回っているため、算出された係数をもとに妥当性について解釈する余地はない。

表 2.9 において、方法の観点から信頼性係数を見ると、自己評価については 1 回目 (M_1) よりも 2 回目 (M_3) の測定の方が信頼性は高くなっており、友人からの評価 (M_2) の信頼性は極端に低かった。信頼性が著しく低い変数に関しては、当然妥当性も低いはずである。実際に、友人からの評価については、 $\text{Con}(x_{ij})$ の値が極端に小さくなっている。したがって、友人からの評価 $x_{12}, x_{22}, \dots, x_{72}$ については、 $\text{MS}(x_{ij})$ と $\text{Con}(t_{ij})$ の値の大きさから妥当性の解釈を行うことは避け、妥当性が低いと結論付けるべきである。

表 2.9 提案モデルの結果 (Steger et al, 2006)

	$\text{Rel}(x_{ij})$	$\text{Con}(x_{ij})$	$\text{MS}(x_{ij})$	$\text{Con}(t_{ij})$
x_{11}	0.680	0.615	0.065	0.904
x_{21}	0.919	0.828	0.091	0.901
x_{31}	0.862	0.829	0.034	0.962
x_{41}	0.864	0.804	0.060	0.931
x_{51}	0.838	0.821	0.018	0.980
x_{61}	0.782	0.767	0.015	0.981
x_{71}	0.709	0.694	0.015	0.979
x_{12}	0.285	0.183	0.102	0.642
x_{22}	0.329	0.257	0.072	0.781
x_{32}	0.498	0.324	0.174	0.651
x_{42}	0.280	0.254	0.026	0.907
x_{52}	0.398	0.223	0.175	0.560
x_{62}	0.331	0.194	0.138	0.586
x_{72}	0.142	0.135	0.007	0.951
x_{13}	0.849	0.816	0.034	0.961
x_{23}	0.989	0.914	0.075	0.924
x_{33}	0.929	0.906	0.023	0.975
x_{43}	0.904	0.824	0.080	0.912
x_{53}	0.881	0.817	0.064	0.927
x_{63}	0.917	0.876	0.041	0.955
x_{73}	0.770	0.721	0.048	0.936

T_1 : 人生の意味の存在 T_2 : 人生の目的 T_3 : 人生への関心指標 T_4 : 人生の満足度
 T_5 : 自尊心 T_6 : 人生の方向付け T_7 : 人生の意味の追求

M_1 : 自己評価 1 回目 M_2 : 友人からの評価 M_3 : 自己評価 2 回目

2.4 研究I 総合考察

本研究の目的は、方法因子の因子得点の和が0になるという仮定の下で、信頼性や妥当性の解釈が1つのデータに対して一通りに定まるようなモデルを提案することであった。

表2.4から表2.6に示したMTMM行列では総じて信頼性係数が高かったのに対して、表2.7のLevin(1973)のMTMM行列では信頼性係数が低かった。信頼性係数が低いということは測定の一貫性が保たれていないことを意味し、当然妥当性も低くなる。Levin(1973)のMTMM行列のように、全体として測定の精度が低い場合には、CT-C(M-1)モデルにおいてどの方法を基準として除くかによる係数の変化が大きいことが確認された。さらに、3種類のモデル間での係数の違いも目立った。他のMTMM行列に比較して、Levin(1973)では基準とする方法を変えることの影響が現れやすかった原因の1つとして、方法の内容に注目する。表2.4のBiesanz & West(2004)では自己評価の繰り返しによる時点差が、表2.5(Coenders & Saris, 2000)と表2.6(Maas et al., 2009)ではそれぞれ尺度やグラフの違いが方法となっているのに対して、Levin(1973)では自己評価、仲間からの評価、オブザーバーからの評価という、評価者の違いを方法としている。この方法の内容の違いが、信頼性係数および妥当性に関する係数の変化に大きく関係していると考えられる。また、表2.8では、自己評価やクラスメイトからの評価と比較してギルフォード検査の信頼性係数の値が大きく、既存の検査が高い信頼性を有していることが示された。Carroll(1952)ではギルフォード検査で仮定されている因子をそのまま特性として用いたために、ギルフォード検査による測定の精度が高められたと考察できる。今回の分析結果から、自己評価の繰り返し測定や、尺度やグラフなど質問紙における提示方法を異にした測定は、安定していることがわかった。一方で、評価者の違いを方法とした測定は信頼性が低くなり、評価者の種類(オブザーバーか、被験者と同等の立場か)によっても測定の信頼性が変化することが示唆された。

CT-CMモデルは、今回の結果からも識別不定や不適解に陥りやすいことが明らかとなった。モデルが識別されず、正常な推定結果が得られる可能性が低いCT-CMモデルを、実際場面で積極的に利用することは難しいだろう。CT-C(M-1)

モデルでは、どの方法を基準とするかによって、 m 通りの信頼性と収束的妥当性、弁別的妥当性に関する解釈が可能となってしまう、特に、信頼性の低いデータに対しては、この影響が深刻な問題となることが示唆された。測定精度の低いデータにCT-C(M-1)モデルを適用した場合、基準とした方法に依存して、信頼性係数や妥当性に関する係数の変化も大きくなり、解釈が一意に定まらない。また、CT-C(M-1)モデルでは、同一のMTMM行列を用いているにもかかわらず、どの方法を基準としてモデルから除くかによって不適解が生じる場合と正常な解が得られる場合があるという例もあった。測定の信頼性や妥当性を検討するにあたって、同一のデータに対して結果が一意に定まらないことは深刻なデメリットである。加えて、本来、CT-C(M-1)モデルの適用は、基準となる方法が意味的に定められるようなデータを対象としており (Eid, Nussbeck, Geiser, Cole, Gollwitzer, & Lischetzke, 2008), この意味でもCT-C(M-1)モデルを利用する際には注意が必要である。

これらのモデルに比較して、方法因子の因子得点の和が0になるという制約を導入した提案モデルは、信頼性および妥当性の解釈を必ず一通りに行うことができ、データを選ばず、より一般的に利用できるモデルである。さらに、本研究で用いた12のMTMM行列に対して提案モデルが識別されない例はなく、19683通りのシミュレーションデータに対してもすべて収束が確認された。したがって、本研究において、提案モデルは信頼性と収束的妥当性、弁別的妥当性に関する解釈がいかなるデータに対しても一意に定まり、かつ、識別の可能性が高い有望なモデルであることが示された。

最後に、今後の課題として、提案モデルの識別に関する数学的証明を挙げる。本研究では、適用例から提案モデルの実用性が確認でき、さらにはシミュレーション研究の結果から本モデルは識別される可能性が非常に高いことが示された。しかし、モデル識別の確認についてはシミュレーション研究だけでは不十分であり、提案モデルの有用性の論拠を示すためにも、識別に関する数学的証明を今後の課題としたい。

第3章 研究II: 特性ごとの合計得点 における信頼性と妥当性の 検討

研究Iの適用例で分析対象とした12種類のMTMM行列について、用いられていた“方法”は次の3種類に大別することができる。1つ目は評価者 (Carroll, 1952; Kelly & Fiske, 1951; Levin, 1973; Manson et al., 2000; Roshal et al., 1971; Steger et al., 2006 の例), 2つ目は測定方法 (Coenders & Saris, 2000; Kothandapani, 1971; Maas et al., 2009; Manson et al., 2000, Muis et al., 2007; Ostrom, 1969 の例), 3つ目は測定機会 (Biesanz & West, 2004; Steger et al., 2006 の例) である。研究Iの結果では, 測定方法や測定機会を“方法”とした場合には, 安定的な測定が行われやすい一方で, 評価者の違いを“方法”とした測定は信頼性が高くなりにくいことが確認された。また, 評価者の立場によって一貫した信頼性や妥当性の傾向が見られたことから, 評価者の立場の違いが測定の信頼性および妥当性に直接的に影響を与える可能性も示唆された。

第3章で論じる研究II, および, 第5章で論じる研究IIIでは, 評価者の違いを“方法”として扱うMTMMデータ, すなわちMTMR (multitrait-multirater) データの分析に焦点を絞って議論を進める。研究IIでは, 複数の評価者による測定値を各被評価者について合計した特性ごとの得点 (特性値) を計算し, その合計得点における信頼性と妥当性を検討する。本章では, はじめに, 現実場面においてMTMR形式のデータが得られるもっとも典型的な例として, 360度フィードバックについて説明を行う。続いて, 360度フィードバックの結果として得られるMTMRデータの信頼性と妥当性に関する研究史を概観し, 自己評価, 上司評価, 同僚評価, 部下評価のそれぞれの特徴を整理する。これらの特徴を踏まえ, 評価者ごとに回答してもらう項目数を変化させた場合の, 合計

得点における信頼性係数と妥当性に関する係数の一般式を導出する。適用例においては、導出した係数を実データに対して計算し、評価者ごとの項目数の配分と、測定の信頼性、収束的妥当性、弁別的妥当性との関係について考察する。最後に、適用例の結果を踏まえ、実際の人事評価場面で本手法を適用することの利点および実用可能性について論じる。

3.1 問題と目的

職場での具体的な行動傾向を把握する方法に、360度フィードバックがある(田中, 2011)。360度フィードバックの最大の特徴は、複数の評価者の眼を通じて、1人の被評価者の資質や行動を評価することにある(Edwards & Ewen, 1996b)。360度フィードバックでは、被評価者の資質や行動がまず自己評価され、続いて、上司・先輩・同僚・部下・後輩・取引先・顧客など、被評価者本人の日頃の仕事ぶりをよく知る評価者による評定が行われる(高橋, 2010)。本人を取り巻く周囲の人材の評価を、貴重なフィードバックデータとして全方位から得ることができるため、この技法は文字通りの意味で360度フィードバックと呼ばれている(Edwards & Ewen, 1996a)。

360度フィードバックの信頼性に関する日本国内の研究としては、入江・鷲坂・舛田・二村(2004, 2005)が挙げられる。入江ら(2004)では、一般化可能性理論を用いることでMOA(M型)(リクルートマネジメントソリューションズ)の分析を行い、対象者、評価者、項目の各要因が測定の信頼性に与える影響の大きさを推定した。信頼性向上のためには、項目数よりも評価者数を増加させることが有効であるというその結果を受けて、入江ら(2005)は項目の相をモデルから除外し、多変量一般化可能性理論(multivariate generalizability theory; Brennan, 2001; Cronbach et al., 1972)の適用により、上司・同僚・部下のそれぞれで、評価者数の変化による一般化可能性の変動を確認した。

入江ら(2004)では、項目数が16という条件の下で0.7以上の信頼性係数を実現するためには上司3~6名、同僚7~8名、部下4~6名の評価者を確保することが望ましいと結論付けている。しかしながら、現実場面では複数の上司や同僚による評定が確保できるとは限らない。通常、上司からの評価は、第一考課

者としての直属管理者1人のみによるものとなる。また、得られる同僚評価の数はその部署の規模に応じて決まるため、小さな部署では同僚が2人以上いないということもあり得る。入江ら(2004)で、上司、同僚、部下いずれにおいても「項目」を含む交互作用の分散成分の割合がもっとも大きくなっていた点からも、複数の評価者が確保できない場合を想定して、各評価者に回答してもらう項目数と、測定の信頼性および妥当性との関連を検討することの重要性が示唆されている。

では、どのような観点から評価者ごとの項目数の配分を決定すれば、より有効なアセスメントを実施できるだろうか。評価者の立場の違いが、測定の信頼性と妥当性に与える影響について着目する。例えば、Conway & Huffcutt (1997)が過去の文献から収集した177の事例をもとにメタ分析を行った結果では、部下評価や同僚評価と比較して、上司評価の信頼性は非常に高いことが示されている。Greguras & Robie (1998)でも、人数と項目数が評価者間で等しいという条件のもとでは、同僚評価や部下評価と比較して上司評価の信頼性が高いと述べられている。一方で、自己評価については、信頼性が低く不正確であるという指摘が様々な研究によってなされている(Ashford, 1989; Bass, 1990; Landy & Farr, 1980)。また、上司と同僚など他者同士の評定値間の相関は高いが、自己評価と他者評価との相関は低くなるという一貫した傾向も数多く報告されている(Harris & Schaubroeck, 1988; Mabe & West, 1982; 二村, 2001; Thornton, 1980)。さらに、妥当性の観点からは、評価者の立場によって重要視する特性が異なるため、上司評価、同僚評価、部下評価はそれぞれ独自の情報を提供し、妥当性の向上に寄与しているという結果がConway, Lombardo, & Sanders (2001)によって明らかになっている。したがって、各評価者に回答してもらう項目数の配分を決定する際に、自己、上司、部下といった評価者の立場の違いによって、特性ごとに測定の精度と正確さへの寄与が異なるという人事アセスメントの視点を活かすことで、より精緻な信頼性と妥当性の検証が可能になると考えられる。

本研究では、360度フィードバックの結果として得られるMTMRデータに対して確認的因子分析モデルを適用し、測定の信頼性および収束的妥当性、弁別的妥当性を検討する方法について論じる。第1章でも述べたように、MTMRデータに確認的因子分析モデルを適用した分析例としては、Conway (1996), Mount et

al. (1998), 高橋 (2001, 2010) が挙げられる。これらの先行研究のうち, Conway (1996) と Mount et al. (1998) では, 個別の MTMR 行列に関して, 信頼性や妥当性の検討までは行われていない。これに対して, 高橋 (2001, 2010) では, 評価者因子のみのモデルを低弁別的妥当性モデル, 特性因子のみのモデルを収束的妥当性モデル, 特性因子と評価者因子を両方仮定したモデルを混合モデルと呼び, それらの適合度を比較することで収束的妥当性と弁別的妥当性に関する考察を行っている。対象者は日本の大手食品会社において実施された管理職研修を受講した本人 177 名と, その上司 316 名, 同僚 255 名, 部下 580 名であった。高橋 (2001, 2010) によると, 低弁別的妥当性モデルの適合度が良い場合には弁別的妥当性が確保されていないことを意味し, 収束的妥当性モデルの適合度が良い場合には収束的妥当性が確保されていることを意味する。また, 混合モデルの適合が良ければ, 弁別的妥当性が部分的に反証されながらも, 収束的妥当性が確保されていると解釈できる, としている。

しかしながら, 高橋 (2001, 2010) の方法もまた, 個々の特性に関する信頼性や妥当性については議論することができない。ここで, 各観測変数について信頼性と妥当性の考察を行うためには, Eid (2000) をもとに研究 I で定義した信頼性係数 $Rel(x_{ij})$, 観測変数の一貫性係数 $Con(x_{ij})$, 観測変数の方法特異性係数 $MS(x_{ij})$ および真値の一貫性係数 $Con(t_{ij})$ を計算し, 結果の解釈を行えばよいだろう。ただし, もともと Campbell & Fiske (1959) が提唱した収束的妥当性と弁別的妥当性の解釈は, 特性 i と方法 j との組み合わせで規定される観測変数 x_{ij} の間で計算される MTMM 相関行列を利用しているため, Eid (2000) による信頼性係数や収束的・弁別的妥当性に関する係数もまた, 変数ごとに算出される指標である。しかし, 360 度フィードバックデータの分析では, それぞれの特性について異なる評価者からの評定結果を合計した得点 (特性値) に興味がある場合も多いだろう。そこで研究 II の目的は, 評価者によって評価項目数を変化させた場合の, 合計得点における信頼性係数と妥当性係数の一般式を導出することとする。

3.2 モデル

研究IIでは、方法因子と特性因子とは無相関とし、特性因子間には互いに相関を仮定するが、方法因子間は無相関とするCT-UMモデルを採用する。すなわち、 Φ_M の対角要素のみ自由母数とし、非対角要素はすべて0に固定するモデルである。因子間相関に対する上述の制約は、評価者の立場の違いを意味する方法因子を系統誤差として扱うという仮定を表現したものであり、評価者ごとの誤差は、同じ立場に属する評価者による測定にのみ寄与し、異なる立場の評価者による測定には寄与しないことを表している (McDonald, 1999)。

信頼性と妥当性の解釈

研究Iで用いた一貫性係数や方法特異性係数は、観測変数の一貫性係数と真値の一貫性係数を比較する必要があるなど、係数の意味付けが複雑で解釈が難しかった。そこで研究IIでは、解釈のしやすさを優先し、研究Iで定義した信頼性や妥当性に関する係数を再定義する。

まず、特性因子と方法因子が無相関であるという仮定により、観測変数の分散は

$$\text{var}(x_{ij}) = \lambda_{T_{ij}}^2 \text{var}(f_{T_i}) + \lambda_{M_{ij}}^2 \text{var}(f_{M_j}) + \text{var}(e_{ij}) \quad (3.1)$$

と表される。(3.1)式右辺の各項は順に、特性因子の影響による変動、方法因子の影響による変動、そしての誤差による変動を表している。

測定の信頼性は、研究Iの場合と同様に、

$$\text{Rel}(x_{ij}) = \frac{\text{var}(T_{ij})}{\text{var}(x_{ij})} + \frac{\text{var}(M_{ij})}{\text{var}(x_{ij})} = \frac{\lambda_{T_{ij}}^2 \text{var}(f_{T_i}) + \lambda_{M_{ij}}^2 \text{var}(f_{M_j})}{\lambda_{T_{ij}}^2 \text{var}(f_{T_i}) + \lambda_{M_{ij}}^2 \text{var}(f_{M_j}) + \text{var}(e_{ij})} \quad (3.2)$$

とする。ここで、(3.2)式の分子は、純粋に特性を反映する部分と、評価者の立場の違いによる系統誤差とに分解される。このとき、

$$\text{Con}(x_{ij}) = \frac{\lambda_{T_{ij}}^2 \text{var}(f_{T_i})}{\lambda_{T_{ij}}^2 \text{var}(f_{T_i}) + \lambda_{M_{ij}}^2 \text{var}(f_{M_j}) + \text{var}(e_{ij})} \quad (3.3)$$

によって、収束的妥当性を検討することが可能である。収束的妥当性は、測定方法が異なっても同一特性の測定間で相関が高いことによって示される。(3.3)

式の収束的妥当性係数は、観測変数全体の変動に対する特性因子の負荷量および分散の相対的な大きさを表しているため、その値が大きいほど、収束的妥当性が高いと解釈できる。

一方で、非弁別的妥当性係数は、

$$\text{Dis}(x_{ij}) = \frac{\lambda_{M_{ij}}^2 \text{var}(f_{M_j})}{\lambda_{T_{ij}}^2 \text{var}(f_{T_i}) + \lambda_{M_{ij}}^2 \text{var}(f_{M_j}) + \text{var}(e_{ij})} \quad (3.4)$$

と定義される。MTMM データにおいて、測定方法が同じであっても、異なる特性を測定していれば、同一の特性を測定した変数同士の場合と比較して相関が低いことによって弁別的妥当性の証拠となる。(3.4)式は、測定値全体の分散に占める系統誤差の影響を表現している。もし系統誤差が大きければ方法の違いが測定値に与える影響が大きいことを意味し、本当に測りたい構成概念と交絡して、特性が違うのに同一方法間での相関が高くなる可能性がある。したがって、系統誤差が小さいと、方法の影響には左右されずに特性の違いが弁別できている状態であり、(3.4)式で算出される値は小さいほど弁別的妥当性が高いと解釈できる。反対に、(3.4)式の値が大きい場合には、特性による変動よりも方法の違いによる変動の方が相対的に大きいことを意味し、弁別的妥当性が低いと結論付けられる。

和得点の信頼性と妥当性

いま、360度フィードバックの実施により、自己評価、上司評価、同僚評価という3つの異なる方法による評価結果が得られているとする。評価者の立場の違いによって測定全体の信頼性と妥当性に与える影響が異なることを考慮して、評価者ごとに評価項目数を変化させた合計得点における信頼性係数と収束的妥当性係数、非弁別的妥当性係数の一般式を導出したい。そこで、自己、上司、同僚に回答してもらう項目数をそれぞれ現状から α_1 倍、 α_2 倍、 α_3 倍して、各評価結果を足し上げた和得点について考える。ただし、自己、上司、同僚のそれぞれで項目数が増減しても、各評価者の立場ごとに、もとの調査票による測定に対してタウ等価測定が成り立っているものとし、かつ、誤差分散の平均は変化しないものと仮定する。

3つの異なる評価者による評価結果を、それぞれ $\{\alpha_1, \alpha_2, \alpha_3\}$ 倍して合計すると、その和得点 $x_{\text{sum}3}$ の分散は

$$\begin{aligned}\text{var}(x_{\text{sum}3}) &= \text{var}(\alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3}) \\ &= (\alpha_1 \lambda_{T_{i1}} + \alpha_2 \lambda_{T_{i2}} + \alpha_3 \lambda_{T_{i3}})^2 \text{var}(f_{T_i}) \\ &\quad + (\alpha_1 \lambda_{M_{i1}})^2 \text{var}(f_{M_1}) + (\alpha_2 \lambda_{M_{i2}})^2 \text{var}(f_{M_2}) + (\alpha_3 \lambda_{M_{i3}})^2 \text{var}(f_{M_3}) \\ &\quad + \alpha_1^2 \text{var}(e_{i1}) + \alpha_2^2 \text{var}(e_{i2}) + \alpha_3^2 \text{var}(e_{i3})\end{aligned}\quad (3.5)$$

となる。 $x_{\text{sum}3}$ を、本研究では再構成スコアと呼ぶ。再構成スコアの分散である(3.5)式を分母として(3.2)式、(3.3)式、(3.4)式を再表現すると、その信頼性係数と収束的妥当性係数、非弁別的妥当性係数はそれぞれ以下のように導出される。

$$\text{Rel}(x_{\text{sum}3}) = \frac{(\alpha_1 \lambda_{T_{i1}} + \alpha_2 \lambda_{T_{i2}} + \alpha_3 \lambda_{T_{i3}})^2 \text{var}(f_{T_i}) + (\alpha_1 \lambda_{M_{i1}})^2 \text{var}(f_{M_1}) + (\alpha_2 \lambda_{M_{i2}})^2 \text{var}(f_{M_2}) + (\alpha_3 \lambda_{M_{i3}})^2 \text{var}(f_{M_3})}{\text{var}(x_{\text{sum}3})}\quad (3.6)$$

$$\text{Con}(x_{\text{sum}3}) = \frac{(\alpha_1 \lambda_{T_{i1}} + \alpha_2 \lambda_{T_{i2}} + \alpha_3 \lambda_{T_{i3}})^2 \text{var}(f_{T_i})}{\text{var}(x_{\text{sum}3})}\quad (3.7)$$

$$\text{Dis}(x_{\text{sum}3}) = \frac{(\alpha_1 \lambda_{M_{i1}})^2 \text{var}(f_{M_1}) + (\alpha_2 \lambda_{M_{i2}})^2 \text{var}(f_{M_2}) + (\alpha_3 \lambda_{M_{i3}})^2 \text{var}(f_{M_3})}{\text{var}(x_{\text{sum}3})}\quad (3.8)$$

なお、(3.5)式から(3.8)式までの方法に関する添字について、1は自己評価、2は上司評価、3は同僚評価に対応している。

3.3 適用例

産業能率大学の人材アセスメントツールの1つである、“ビジネス基礎力診断S-BASE”による360度フィードバックの結果を用いて適用例を示す。

概要 ビジネス基礎力診断では、企業における若手社員から中堅社員を対象者として、以下の3つの領域でビジネスに必要な基礎的能力¹を設定し、測定している(学校法人産業能率大学総合研究所, 2012)。各領域の内容と、該当する項目数を以下に示した。

¹本調査では、3つの“ビジネスに必要な基礎的能力”が測定したい構成概念であり、それぞれMTMMデータにおける“特性”として扱われる。

- “自己確立”領域：自分自身の確固とした想いや考えを持って、常に成長を目指す力。(34項目)
- “仕事確立”領域：仕事のサイクルに従って堅実に業務を進め、目標を達成する力。(33項目)
- “関係性確立”領域：影響力を発揮し、周囲と協働を図りながら、組織としての成長を出す力。(31項目)

調査 内容および個人情報秘匿されることを説明した上で、計98項目から構成される調査票（質問紙）を用いて、“自己評価”、“上司評価”、“同僚評価”による回答を求めた。“上司評価”は各被評価者に対して上司1人による評定であり、“同僚評価”は、1人の被評価者に対して、1人～最大6人までの同僚による評定である。

質問項目への反応は“あてはまらない”、“ややあてはまらない”、“ややあてはまる”、“あてはまる”の4件法とした。また他者(上司・同僚)評価調査票の項目は、自己評価調査票と同一内容としたが、文章表現を“その人は～”で始まる形式にした。

期間 2010年12月から2011年2月までの3ヶ月間。

回収方法 調査票の受け渡しは郵送で行った。

対象 日本に本社を置く31社の企業。企業の所在地は関東地方21社、中部地方4社、近畿地方3社、中国地方3社であり、業種は農林、住宅、食品、紙・パルプ、化学・薬品、機械・電気、精密、商社、小売・飲食、運輸、サービス、教育、医療など多岐にわたった。また、調査対象者は入社2年目から10年目までの社員とし、“自己評価”については855人（男性641人、女性211人、不明3人），“上司評価”については446人，“同僚評価”については1369人の回答を得た。

分析にあたって、領域ごとに“自己評価”、“上司評価”、“同僚評価”の合計得点を算出した。この際、“同僚評価”については、あらかじめ被評価者ごとに同

僚の人数に応じて平均値を計算した。その結果、すべての領域で欠損値のなかった有効回答数は 401（平均 25.8 歳，SD3.55）となった。

結果と考察

ビジネス基礎力診断データの相関行列を表 3.1 に示した。相関行列を 3×3 のブロック行列に分割して観察すると，“自己評価”と“上司評価”，および，“自己評価”と“同僚評価”の相関は総じて低いのに対して，“上司評価”と“同僚評価”との相関の方が高い。同一特性・異方法の相関係数の平均値は 0.347，異特性・異方法の相関係数の平均値は 0.296，異特性・同一方法の相関係数の平均値が 0.832 であった。同一方法，すなわち同じ立場の評価者であれば，特性が異なっても相関が非常に高いことが示された。一方で，評価者の立場が異なると，特性の異同にかかわらず一貫して相関係数が低い傾向にある。したがって，収束的妥当性および弁別的妥当性の積極的な証拠は得られなかった。しかし，“自己評価”と他者評価との相関は総じて低いこと，“自己評価”に比較して他者評価の方が妥当性が高い傾向にあることなど，二村(2001)で紹介されている 360 度フィードバックデータの一般的な特徴と一致した傾向が見られた。

表 3.1 ビジネス基礎力診断データの相関行列

	自己評価			上司評価			同僚評価			
	自己	仕事	関係性	自己	仕事	関係性	自己	仕事	関係性	
自己評価	自己	1.00								
	仕事	0.79	1.00							
	関係性	0.72	0.73	1.00						
上司評価	自己	0.28	0.26	0.23	1.00					
	仕事	0.21	0.29	0.19	0.89	1.00				
	関係性	0.21	0.24	0.28	0.86	0.84	1.00			
同僚評価	自己	0.28	0.22	0.20	0.47	0.46	0.43	1.00		
	仕事	0.27	0.28	0.24	0.46	0.51	0.46	0.92	1.00	
	関係性	0.22	0.18	0.26	0.42	0.41	0.47	0.88	0.86	1.00

データの概要で述べたように，“ビジネス基礎力診断 S-BASE”では自己，上司，同僚のすべての評定を通して，特性ごとに決まった項目数で実施されていた。以下では，(3.6)，(3.7)，(3.8)式を用いて，評価者ごとの項目数と，再構成スコアの信頼性・妥当性との関係を考察する。まず，特性因子間にのみ相関を仮定したCT-UMモデルを適用し，確認的因子分析を実行した結果，適合度指標の値は，CFI=0.994，TLI=0.985，RMSEA=0.063であった($\chi^2=38.821$ ， $df=15$)²。推定されたCT-UMモデルのパラメタの非標準化解は，表3.2の通りである³。

表 3.2 CT-UM モデルによる推定結果と信頼性係数および妥当性係数

変数	λ_T	λ_M	誤差分散	Rel(x_{ij})	Con(x_{ij})	Dis(x_{ij})
x_{11}	1.000	1.000	34.419	0.815	0.256	0.559
x_{21}	1.000	0.937	22.832	0.852	0.262	0.591
x_{31}	1.000	0.839	36.800	0.751	0.256	0.495
x_{12}	1.318	1.000	12.594	0.936	0.420	0.516
x_{22}	1.521	1.017	21.268	0.903	0.426	0.477
x_{32}	1.451	0.862	18.542	0.893	0.459	0.435
x_{13}	1.116	1.000	4.246	0.971	0.408	0.563
x_{23}	1.374	0.948	9.706	0.939	0.478	0.461
x_{33}	1.097	0.863	12.474	0.895	0.382	0.513
	特性因子間の分散共分散行列			方法因子間の分散共分散行列		
f_1	47.667		<i>sym.</i>	104.063	0	0
f_2	40.213	40.494		0	101.650	0
f_3	36.738	33.619	37.899	0	0	82.049

表 3.2 には，モデルの母数の推定値および，因子負荷量と因子の分散，誤差分散の値を(3.2)式から(3.4)式に当てはめて，変数ごとに信頼性係数と収束的妥当性係数，非弁別的妥当性係数を算出した結果も併せて示した。表 3.1 の相関係数の観察結果と同様に，“自己評価”に比較して，“上司評価”と“同僚評価”

²分析には，統計解析環境 R のパッケージ lavaan を用いた。

³表 3.1 では Campbell & Fiske (1959) に従い，信頼性および収束的妥当性，弁別的妥当性の解釈を行うために相関行列を示したが，信頼性係数と妥当性に関する係数を計算する際には，生データをもとに推定された CT-UM モデルの非標準化解を用いた。特に，再構成スコアの信頼性と妥当性を考えるとき，分析に相関行列を用いたり，係数の計算時に標準化解を用いたりすると，データを基準化した上で項目数の増減を検討していることになるため，本手法の適用にあたっては，共分散行列，もしくは生データから計算される非標準化解を利用する必要がある。

では信頼性、収束的・弁別的妥当性ともに高いことが確認できた。

係数の算出

表 3.2 の推定値を用いて、(3.6)，(3.7)，(3.8) 式に従って，評価者ごとに項目数を変化させた再構成スコアの信頼性係数と妥当性に関する係数を算出した。“自己評価”，“上司評価”，“同僚評価”それぞれの項目配分 $\{\alpha_1, \alpha_2, \alpha_3\}$ を変化させたときの，再構成スコアの信頼性係数 $\text{Rel}(x_{\text{sum}3})$ および収束的妥当性係数 $\text{Con}(x_{\text{sum}3})$ ，非弁別的妥当性係数 $\text{Dis}(x_{\text{sum}3})$ の変動を表 3.3 に示した。

表 3.3 項目数の違いによる係数の変動

$\alpha_1:\alpha_2:\alpha_3$	1:1:1	1:1.5:0.5	1:0.5:1.5	1.5:1:0.5	1.5:0.5:1	
特性 1 の $\text{Rel}(x_{\text{sum}3})$	0.943	0.937	0.947	0.908	0.910	
特性 1 の $\text{Con}(x_{\text{sum}3})$	0.624	0.588	0.594	0.548	0.545	
特性 1 の $\text{Dis}(x_{\text{sum}3})$	0.319	0.349	0.353	0.360	0.364	
特性 2 の $\text{Rel}(x_{\text{sum}3})$	0.943	0.931	0.946	0.922	0.927	
特性 2 の $\text{Con}(x_{\text{sum}3})$	0.655	0.603	0.639	0.579	0.590	
特性 2 の $\text{Dis}(x_{\text{sum}3})$	0.288	0.328	0.307	0.343	0.337	
特性 3 の $\text{Rel}(x_{\text{sum}3})$	0.910	0.906	0.905	0.873	0.869	
特性 3 の $\text{Con}(x_{\text{sum}3})$	0.632	0.607	0.590	0.563	0.548	
特性 3 の $\text{Dis}(x_{\text{sum}3})$	0.278	0.399	0.315	0.310	0.321	
$\alpha_1:\alpha_2:\alpha_3$	0.5:1:1.5	0.5:1.5:1	1.25:1.25:0.5	1.25:0.5:1.25	0.5:1.25:1.25	
特性 1 の $\text{Rel}(x_{\text{sum}3})$	0.967	0.959	0.924	0.929	0.963	
特性 1 の $\text{Con}(x_{\text{sum}3})$	0.629	0.619	0.577	0.578	0.632	
特性 1 の $\text{Dis}(x_{\text{sum}3})$	0.338	0.339	0.347	0.351	0.331	
特性 2 の $\text{Rel}(x_{\text{sum}3})$	0.952	0.942	0.928	0.938	0.948	
特性 2 の $\text{Con}(x_{\text{sum}3})$	0.663	0.638	0.601	0.623	0.659	
特性 2 の $\text{Dis}(x_{\text{sum}3})$	0.289	0.304	0.328	0.315	0.290	
特性 3 の $\text{Rel}(x_{\text{sum}3})$	0.928	0.926	0.892	0.889	0.928	
特性 3 の $\text{Con}(x_{\text{sum}3})$	0.630	0.633	0.594	0.577	0.640	
特性 3 の $\text{Dis}(x_{\text{sum}3})$	0.298	0.293	0.298	0.311	0.289	
$\alpha_1:\alpha_2:\alpha_3$	0.5:1:1	1:0.5:1	1:1:0.5	1:0.5:0.5	0.5:1:0.5	0.5:0.5:1
特性 1 の $\text{Rel}(x_{\text{sum}3})$	0.961	0.933	0.928	0.909	0.949	0.960
特性 1 の $\text{Con}(x_{\text{sum}3})$	0.636	0.592	0.590	0.554	0.612	0.623
特性 1 の $\text{Dis}(x_{\text{sum}3})$	0.325	0.341	0.338	0.355	0.337	0.337
特性 2 の $\text{Rel}(x_{\text{sum}3})$	0.948	0.940	0.931	0.925	0.938	0.951
特性 2 の $\text{Con}(x_{\text{sum}3})$	0.662	0.634	0.615	0.593	0.628	0.662
特性 2 の $\text{Dis}(x_{\text{sum}3})$	0.286	0.305	0.317	0.333	0.310	0.289
特性 3 の $\text{Rel}(x_{\text{sum}3})$	0.927	0.894	0.896	0.872	0.918	0.920
特性 3 の $\text{Con}(x_{\text{sum}3})$	0.643	0.593	0.605	0.563	0.628	0.621
特性 3 の $\text{Dis}(x_{\text{sum}3})$	0.283	0.301	0.291	0.309	0.290	0.299

配分が $\{\alpha_1 : \alpha_2 : \alpha_3\} = \{1 : 1 : 1\}$ のときは，項目数を現状維持で，手元のデータにおいて特性ごとの単純な和得点を考えたときの信頼性と妥当性を意味す

るため、ベースラインとして参照することができる。 $\{\alpha_1 : \alpha_2 : \alpha_3\} = \{1 : 1 : 1\}$ のとき、3つの特性ともに信頼性係数は0.9以上と十分に大きな値であり、妥当性係数の値も特性間でほとんど違いはなかった。

S-BASEは、“自己確立”34項目、“仕事確立”33項目、“関係性確立”31項目の計98項目から構成されており、現状でも回答者への負担が大きくないとは言いがたい。360度フィードバックを現場で実施する場合には、実施時間の制約もあるだろう。そこで、実施時間を短縮せざるを得ない場合に、 $\{\alpha_1 : \alpha_2 : \alpha_3\}$ を変化させ、係数を計算することで、自己、上司、同僚のいずれの項目をどれくらい減らすと信頼性や妥当性がどの程度低下するのか、あるいは低下しないのかを事前に予測できることは、実践的意義が大きいと考えられる。

以下では、回答者への負担軽減および実施時間短縮を目的とする実践場面を想定して、(1)項目数の合計が調査票全体で変わらない場合と、(2)項目数を現状維持または減らす場合の2パターンに分けて、 $\{\alpha_1 : \alpha_2 : \alpha_3\}$ を変化させ、再構成スコアの信頼性と妥当性がどのように変化するか観察する。項目配分 α を検討することで、3つの評価者の立場による測定全体(合計得点)としての信頼性や妥当性をできるだけ高く維持した状態で、効率よく項目数を減らすためには、どの立場からの評定をより重視すべきかという示唆を得ることができる。

(1) 項目数の合計が調査票全体で変わらない場合

表3.3の1段目と2段目はいずれも、 $\alpha_1 + \alpha_2 + \alpha_3 = 3$ となるような α の組み合わせである。これは、評価者の立場ごとに調査票の項目数を変化させるけれども、全体としては項目数の合計が変化しない状況を意味する。

$\alpha_1 : \alpha_2 : \alpha_3 = \{1 : 1.5 : 0.5\}$ と $\alpha_1 : \alpha_2 : \alpha_3 = \{1 : 0.5 : 1.5\}$ の結果から、“自己評価”の項目数は現状のままでも、“上司評価”か“同僚評価”のいずれかを減らすと、ベースラインと比較して信頼性係数の値にはあまり変化がないものの、妥当性係数に与える影響は大きいことがわかる。特に、特性3“関係性確立”の弁別的妥当性の低下が顕著である。

$\alpha_1 : \alpha_2 : \alpha_3 = \{1.5 : 1 : 0.5\}, \{1.5 : 0.5 : 1\}, \{1.25 : 1.25 : 0.5\}, \{1.25 : 0.5 : 1.25\}$ のように、“自己評価”の項目数を増やすと、妥当性ばかりでなく信頼性係数の低下にもつながる。反対に、 $\alpha_1 : \alpha_2 : \alpha_3 = \{0.5 : 1 : 1.5\}, \{0.5 : 1.5 : 1\}, \{0.5 : 1.25 : 1.25\}$ のように“自己評価”の項目数を減らして、“上司評価”と

“同僚評価”の一方、あるいは両方の項目数を増やすと、総じて信頼性、妥当性ともにベースラインよりも向上した。

また、信頼性に対する項目数 $\{\alpha_1 : \alpha_2 : \alpha_3\}$ の変化の影響をもっとも受けにくいのは特性2の“仕事確立”であった。これに対して、“自己確立”と“関係性確立”の信頼性は、評価者ごとの項目数の変化に敏感である。“ビジネス基礎力診断 S-BASE”で測定の対象としている3つの特性のなかでは、“仕事確立”がもっとも客観的な判断がしやすい領域であることから、項目数の増減に左右されず、評価者の立場が違ってても一貫して安定的な測定がなされやすかったのかもしれない。

項目数の合計が調査票全体で変わらない場合という条件の下で、最適な項目数の配分について検討すると、 $\{\alpha_1 : \alpha_2 : \alpha_3\} = \{0.5 : 1 : 1.5\}$ のときに信頼性係数はもっとも大きな値となった。ただし、収束的妥当性係数および非弁別的妥当性係数の値も含め総合的に判断すると、信頼性と妥当性の両方が改善されるのは $\alpha_1 : \alpha_2 : \alpha_3 = \{0.5 : 1.25 : 1.25\}$ の場合であった。

(2) 項目数を現状維持または減らす場合

表3.3の最下段は、 α の値をいずれも1あるいは0.5とし、項目数を現状維持、あるいは減らす場合のみを考えている。これは、回答時間や回答者の負担、同一特性を測定すると見なせる項目数が限られているといった理由から、現状より項目を増やすことが難しい状況を想定している。

この場合でも、“上司評価”と“同僚評価”のいずれかでも項目数を少なくすると、信頼性、妥当性ともにベースラインよりも低くなることがわかる。一方で、“自己評価”のみ項目数を半分にして、 $\alpha_1 : \alpha_2 : \alpha_3 = \{0.5 : 1 : 1\}$ とした場合には、ベースラインに比較していずれの特性でも信頼性と収束的妥当性が改善する傾向が見られた。

(1) で見た $\{\alpha_1 : \alpha_2 : \alpha_3\} = \{1.5 : 0.5 : 1\}$ と、(2) の $\{\alpha_1 : \alpha_2 : \alpha_3\} = \{0.5 : 0.5 : 1\}$ を比較すると、他者評価の項目配分は両者で同じであるにもかかわらず、“自己評価”の項目数を半分にした(2)の場合の方が、信頼性および収束的妥当性と弁別的妥当性が向上している。(1)、(2)両方の場合を通して、“自己評価”は測定の信頼性と妥当性の維持および改善にはあまり寄与しないことが明らかとなった。(2)の結果より、項目数を増やすことができない場合には、“自

己評価”の項目数を減らすことで、測定の信頼性と妥当性を維持しつつ、より効率的な360度フィードバックを行える可能性が示唆された。

一般に、“自己評価”の項目を減らすと、本人の自己認識が不確かになってしまい、行動変容が起きにくくなるという懸念も考え得る。しかし、“自己評価”を減らしても、測定の信頼性・妥当性が低下しないということは、測定したい特性に関して、他者評価の方がより正確かつ安定的に測定できているということの意味する。これは、“自己評価”の項目数を減らしても情報損失は最小限に抑えることができると保証されているということである。したがって、“自己評価”の項目数を減らしても、他者からのフィードバックが充実していれば、その結果を考察することで自己の気づきを促し、行動変容に結びつけることができると考えられる。

ただし、実際に評価者間で回答してもらう項目数を変えて360度フィードバックを実施する場合には、項目選定および被評価者に結果をフィードバックする際に工夫が必要となる。例えば、(1)の結果に従い、“自己評価”の項目数を減らして“上司評価”と“同僚評価”の項目を増やすとき、“自己評価”に使用した項目のみがすべての立場で共通した項目となる。追加する項目については、内的整合性を考慮して、調査票全体としての因子構造が保たれるよう留意して作成しなくてはならない。一方で、“自己評価”の項目を減らすとき、項目内容を十分に吟味した上で、経験的に重要と考えられる項目を残すように、かつ下位領域をまんべんなく網羅するように選定すべきである。このような工夫の下では、評価者間の共通項目のみを利用することによっても、有意義なフィードバックが可能となるであろう。

係数ごとの特徴

以下では、 α_1 の値を固定した上で、 α_2 と α_3 をそれぞれ0.25から2.0までの範囲で変化させたときの各係数の変動について考察する。

特性3の信頼性に関する考察 まず、ベースラインでもっとも信頼性係数が低かった特性3“関係性確立”に注目して、信頼性係数 $\text{Rel}(x_{\text{sum}3})$ の変動を等高線によって表したのが図3.1の上段である。左から順に、 $\alpha_1 = 0.5, 1.0, 1.5$ の場合

を示している。これら3つの図から、信頼性の向上に対する“上司評価”と“同僚評価”の影響力はほぼ同じであることが見てとれる。

“上司評価”および“同僚評価”の項目数を増やすと、信頼性は向上する傾向があるものの、同時に“自己評価”の項目数も増やすと、“上司評価”や“同僚評価”の項目数を増やしても信頼性の改善傾向が鈍ることがわかった。“自己評価”の項目数は現状維持以下とすることを前提とすれば、“上司評価”と“同僚評価”の項目数をそれぞれ現状から1.5倍にすることで、信頼性係数を0.92以上に改善できる可能性が示唆された。

特性1の収束的妥当性に関する考察 次に、ベースライン時にもっとも $\text{Con}(x_{\text{sum}3})$ の値が小さかった特性1“自己確立”について、項目数の変化に伴う収束的妥当性係数の変動を考察する。図3.1の中段では、それぞれ $\alpha_1 = 0.5, 1.0, 1.5$ としたときの再構成スコアの収束的妥当性係数を、左側から順に図示した。左側から右側へ“自己評価”の項目数 α_1 を増やすにしたがって、等高線の間隔は広く、カーブは緩やかになっている。

図3.1の中段左図より、 $\alpha_1 = 0.5$ と“自己評価”の項目を現状より減らした場合でも、“上司評価”と“同僚評価”が $\alpha_2 = \alpha_3 = 1$ と現状維持であれば、収束的妥当性係数がベースラインの0.624を下回ることはないことがわかる。それどころか、 $\alpha_1 = 0.5$ のとき、 α_2 と α_3 を小さくして、それぞれ0.75としてもなお、収束的妥当性係数は $\text{Con}(x_{\text{sum}3}) = 0.63$ の領域に留まっている。

α_2 の値を固定して、 α_3 の値を大きくしていくと、ある一定の値までは収束的妥当性係数も大きくなるものの、それ以降は α_3 をいくら大きくしても収束的妥当性係数が再び小さくなってしまう様子を図3.1の中段左図は表している。これは α_3 を固定して、 α_2 を大きくする状況でも同じである。例えば、 $\alpha_2 = 1$ とすると、 α_3 が約0.7以上で $\text{Con}(x_{\text{sum}3})$ の値は0.63を上回るが、さらに α_3 を1.5より大きくすると、収束的妥当性係数は再び小さくなってしまう。これに対して中図と右図では等高線のカーブが緩やかになっており、 α_1 が1以上のときには、 α_2 と α_3 のいずれか一方のみを増加させた場合でも、ある一定の項目数以上にすると急激に収束的妥当性が低下する、という現象は起こらないことを意味している。

また、図 3.1 中段右図から、 α_1 を大きくすると、 α_2 および α_3 をどんなに大きくしても、収束的妥当性係数の値をベースラインの値である 0.624 より大きくすることはできないことが示された。例えば、 $\alpha_1 = 1.5$ と“自己評価”の項目数のみ増やし、“上司評価”と“同僚評価”については $\alpha_2 = \alpha_3 = 0.5$ とすると、 $\{\alpha_1 : \alpha_2 : \alpha_3\} = 1 : 1 : 1$ のときに 0.624 だった収束的妥当性係数は約 0.5 まで低下することがわかる。

特性 1 の弁別的妥当性に関する考察 最後に、非弁別的妥当性係数の変動については、 $\text{Dis}(x_{\text{sum}3})$ の値がベースラインで 0.3 を上回っていた特性 1 を考察の対象とする。図 3.1 下段の 3 つの図の比較から、 α_1 の値を大きくするにつれて非弁別的妥当性係数が大きくなり、弁別的妥当性が低くなることが示された。

図 3.1 下段中図より、“自己評価”の項目数は変えず $\alpha_1 = 1$ としたときに、もともと弁別的妥当性を高くするには、 $\alpha_2 = \alpha_3 = 1.0 \sim 1.5$ とすればよいことがわかる。この領域から α_2 と α_3 を大きくしても小さくしても、非弁別的妥当性係数の値は大きくなってしまふ。

$\alpha_1 = 1.5$ のときについて図示した図 3.1 右図は、“上司評価”の項目配分 α_2 と“同僚評価”の項目配分 α_3 をそれぞれ大きくするにしたがって非弁別的妥当性係数が小さな値となることを表しているが、 $\alpha_2 = \alpha_3 = 2$ としてもなお 0.32 であり、ベースラインの 0.319 を上回っている。すなわち、もともと測定精度の低い“自己評価”を増やしたことで、他者評価の項目数を増やしても弁別的妥当性が改善されにくくなっていると解釈できる。

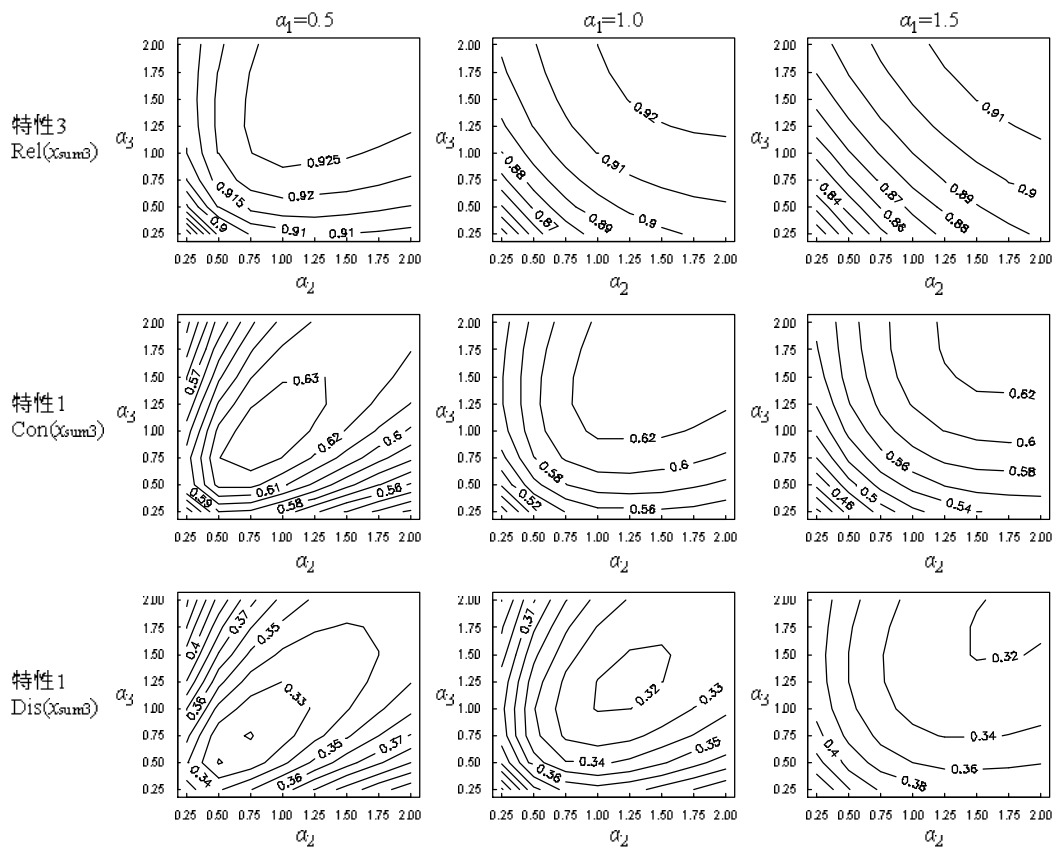


図 3.1 項目数の変化に伴う再構成スコアの信頼性係数と妥当性係数の変動

3.4 研究Ⅱ 総合考察

研究Ⅱの目的は、評価者によって調査票の評価項目数を変化させた場合の、再構成スコアにおける信頼性係数と妥当性係数の一般式を導出し、360度フィードバックにおける項目数と、測定の信頼性および収束的妥当性、弁別的妥当性との関係を考察することであった。360度フィードバックデータの信頼性と妥当性を改善するためには、もちろん他者評価に関して評価者の人数を増やすことも有効である。しかしながら、日本企業においては、上司評価は一次人事考課者に該当する1人のみが担当することが多く、収集可能な同僚評価の数も部署の規模に応じて決まってくる。したがって、評価者間での項目数の配分と、測定の信頼性、妥当性との関係を明確にすることは、360度フィードバックの適正な運用のための有益な示唆となり得る。

360度フィードバックデータをMTMMデータとして扱い、確認的因子分析モデルを適用することで、再構成スコアの信頼性係数と収束的妥当性係数、非弁別的妥当性係数を定義した。これらの係数を、360度フィードバックのためのツール“ビジネス基礎力診断 S-BASE”の実施結果に対して計算した結果から、評価者ごとに調査票の項目数を変化させ、適切に配分することで、より効率的に測定の信頼性と妥当性を高められることが確認できた。分析の結果、“自己評価”の項目数を増やすと信頼性と妥当性は低下し、一方で、“上司評価”と“同僚評価”の項目数を増やすことは信頼性と妥当性の向上に貢献することが示された。これらの傾向は、本データにおいては特に収束的妥当性と弁別的妥当性に関して顕著であった。今回用いた“ビジネス基礎力診断 S-BASE”で、現実的に実行可能な範囲内で項目数を変更するとすれば、“自己評価”については半分に減らして約17項目とし、上司と同僚に回答してもらう項目数は現状の1.25倍の40項目前後まで増やすことで、いずれの特性においても全体的に信頼性と妥当性が向上する可能性が示唆された。

人事考課に携わるのは上司であるという事実に鑑みても、上司に回答してもらう項目数を増やして測定の信頼性と妥当性を高めようという試みは合理的である。また、今回の分析結果において、“上司評価”と同様に、“同僚評価”の項目数の増加もまた、信頼性と妥当性の向上につながるということが明らかとなった。

この結果は、自己評価と人事評価との相関はほとんどないのに対して、同僚評価を含む他者評価は人事評価と比較的高い相関があるという二村(1998)の報告を支持するものである。360度フィードバックを利用してより適切な人事アセスメントを実施するためには、人事考課に直接関わる上司だけではなく、同僚評価も疎かにしてはならない。

現実の360度フィードバック実施場面では、項目数を減らすことは容易でも、増やすことは困難な場合が多いであろう。回答者への負担、あるいは実施に要する時間やコスト等の制約により、現状よりも項目数を増やすことが難しい場合には、“自己評価”の項目数を減らすことによってもまた、信頼性と妥当性の向上が見込まれる。再構成スコアの信頼性係数と(非)妥当性係数は、目標とする信頼性係数と収束的妥当性係数、非弁別的妥当性係数を設定した上で、現状から項目数をどれくらい減らしてもその目標を達成できるかについて検討するためにも有用である。測定の質は高い水準で保ちつつ、各評価者への負担を必要最小限に抑えた理想的な人事アセスメントツールの作成のために、本手法を役立てることができるだろう。

以上より、(3.6)、(3.7)、(3.8)式を利用することで、測定全体の信頼性と妥当性という観点から、実現可能な範囲内で、評価者ごとに最適な項目数の配分を検討できる可能性が示唆された。一方で、本研究の結果から得られる項目配分に関する示唆は、信頼性と妥当性という心理測定論的観点のみに基づく判断であり、項目配分 α を変化させることの実践的意義を直接明らかにすることは難しいという課題も残る。信頼性や妥当性が極端に損なわれない範囲で項目数を減らすことが可能なのであれば、フィードバックの際の情報損失も最小限に抑えられると見なすことができ、実施時間短縮や評価者への負担軽減といったメリットの方が大きいと考えられる。しかしながら、評価者の立場によって回答してもらった項目数を変更することが、実践場面でどのような意味を持つのかについては、今後慎重に検討する必要があるだろう。

第4章 ベイズ統計学とマルコフ連鎖 モンテカルロ法

本章では、研究 III において用られるハミルトニアンモンテカルロ (Hamiltonian Monte Carlo; HMC; Duane, Kennedy, Pendleton & Roweth, 1987, Neal, 1996) 法について理解を深めるため、ベイズ統計学および MCMC 法とその推定アルゴリズムについて概説する。

4.1 ベイズ統計学

ベイズ統計学 (Bayesian statistics) は、以下に述べるベイズの定理 (Bayes' theorem) に基づいて推論を行う統計的推測のアプローチである。

ベイズの定理

事象 B が与えられた下で、事象 A が観察される確率は

$$\Pr(A|B) = \frac{\Pr(B|A) \times \Pr(A)}{\Pr(B)} \quad (4.1)$$

のように表現され、(4.1) 式はベイズの定理と呼ばれる。この定理は、イギリスの長老派の牧師トーマス・ベイズ (Thomas Bayes, 1702-1761) にその名の由来を持つ。ベイズの定理は、 A が仮説で、 B が実際に得られる結果であるとして解釈することが可能である。 $\Pr(A)$ は実際に結果が得られる前の、仮説 A に対する事前情報、あるいは確信の度合いを表す主観的な確率であり、事前確率 (prior probability) と呼ばれる。これに対して、 $\Pr(A|B)$ は事後確率と呼ばれ、実際に結果 B が手元に得られた後に更新された、仮説 A についての確信の強さを表す主観確率である。また、 $\Pr(B|A)$ は仮説 A の下で結果 B が実際に発生する確率であるが、結果が生じてしまった後ではもはや確率ではなくなるため、尤度と

呼ばれる。 $\Pr(B)$ は $\Pr(B|A)$ を周辺化した後の、結果 B が得られる確率を表している。

(4.1)式は、個別の事象から、確率変数および分布を扱う場合へと拡張することが可能である。確率変数 \mathbf{x} と \mathbf{x} に関する未知のパラメタ $\boldsymbol{\theta}$ に関して、ベイズの定理は

$$\Pr(\boldsymbol{\theta}|\mathbf{x}) = \frac{\Pr(\mathbf{x}|\boldsymbol{\theta}) \times \Pr(\boldsymbol{\theta})}{\Pr(\mathbf{x})} \quad (4.2)$$

と表現される。このとき(4.2)式の $\Pr(\boldsymbol{\theta})$ はデータが得られる以前の $\boldsymbol{\theta}$ が従っている分布を表し、事前分布 (prior distribution) と呼ばれる。また、 $\Pr(\mathbf{x}|\boldsymbol{\theta})$ はパラメタ $\boldsymbol{\theta}$ に関する尤度であり、 $\Pr(\boldsymbol{\theta}|\mathbf{x})$ は実際にデータが得られた後の $\boldsymbol{\theta}$ についての事後分布を表している。したがって、(4.2)式は、事前分布がデータを通じて事後分布へと変化する仕方を示す式となっている (渡部, 1999)。

(4.2)式に従って、事前分布からデータがもたらす情報を反映した事後分布を作り出すために、ベイズの定理を用いてパラメタの推論を行おうとする統計学をベイズ統計学という。しかし、ベイズ統計に基づく母数推定を実際に行うことは、必ずしも容易ではない。 $\Pr(\mathbf{x})$ はデータ \mathbf{x} の分布を表すものであるが、これを定めることは難しいため、

$$\Pr(\mathbf{x}) = \int \Pr(\mathbf{x}|\boldsymbol{\theta}) \Pr(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (4.3)$$

という形で、尤度関数から母数 $\boldsymbol{\theta}$ を積分消去することによって求めるのが一般的である。このように、ベイズの定理の右辺では、高次元の確率分布に関する積分を必要とする。関数の形状が複雑な場合には、積分操作を代数的に行うことがほとんど不可能であり、数値積分による近似を利用せざるを得ない。ところがこの高次元関数の数値積分は莫大な計算量を要するため、現実的には利用することが困難である。

また、パラメタが多い場合には、最終的な結論である事後分布も高次元の関数となるため、これを解釈することも難しい。事後分布から興味の対象ではない母数を積分消去することによって、個々のパラメタについて解釈を行うことが可能となるものの、この積分消去の計算もまた莫大な計算量となる。

これらの理由により、ベイズ推測が利用されるのは、扱う統計モデルが単純で、(4.2)式の右辺が代数的に解ける場合だけにとどまっていた。また、そのような場

合であっても、できるだけ計算を簡単にするための工夫として、共役 (conjugate) な事前分布の利用が欠かせなかった。共役な事前分布とは、ある尤度関数に対して、事前分布と事後分布が同じ分布族になるような事前分布のことである。事前分布が共役であれば、事後分布が扱いやすい形になることが保証されるため、従来のベイズ推定においてはこのような事前分布を用いることが多かった。しかし、事前分布は本来、母数の値に対する事前知識を反映させるためのものであり、計算上の困難さの軽減という便宜的な要請により制約されるのは望ましいことではない。こういった制限のために、これまでベイズ推測の枠組みが広く用いられることはなかった。

この状況は、コンピュータの計算能力の向上と利用コストの低下に伴い、マルコフ連鎖モンテカルロ法と呼ばれるサンプリング法が手軽に利用可能となったことで、大きく変化することとなった。MCMC法を用いれば、高次元の数値積分を実行せずとも、事後分布からの疑似標本によって構成されたマルコフ連鎖を構築することができる。また、仮にパラメタの数が多い場合でも、特定のパラメタに関して、その他の母数による条件付事後分布を目標分布とすることで、同様に数値積分を避けることができる。

4.2 マルコフ連鎖モンテカルロ法

本節では、豊田 (2008) に基づき、マルコフ連鎖モンテカルロ法について概説する。

マルコフ連鎖

時点を $t = 1, 2, \dots, T$ とするとき、各時点における状態を実現値とする確率変数を $x^{(t)}$ と表す。このように時間に従って確率的に状態が変化する系列を確率過程という。また、 $x^{(t)}$ のとりうるすべての状態を含む集合を状態空間 (state space) と呼び、 S で表す。

ここで、時点 t において状態が $i (i \in S)$ にある事象が、次の $t+1$ 時点において j となる確率は、過去の確率過程の実現値が与えられた下での $x^{(t+1)}$ の条件

付確率として、次のように表現される。

$$P(x^{(t+1)} = j | x^{(1)} = i_1, \dots, x^{(t)} = i) \quad (4.4)$$

この条件付確率 (4.4) 式が、すべての t について

$$P(x^{(t+1)} = j | x^{(1)} = i_1, \dots, x^{(t)} = i) = P(x^{(t+1)} = j | x^{(t)} = i) \quad (4.5)$$

という性質を満たすならば、その確率過程はマルコフ連鎖 (markov chain) と呼ばれる。(4.5) 式は、任意の時点 t において、 $x^{(t)}$ が状態 i にあるとき、次の時点 $t+1$ において $x^{(t+1)}$ が状態 j へと推移する確率は、現在の状態 $x^{(t)}$ のみに依存することを表している。したがって、マルコフ連鎖では、 $x^{(t)}$ が所与の下では、 $t-1$ 以前の時点の状態 $x^{(1)}, \dots, x^{(t-1)}$ とは独立となる。

(4.5) 式で示されたように、マルコフ連鎖は現在の状態のみが次の状態に影響する確率過程の系列である。ここで、任意の時点 t から次の時点 $t+1$ に推移する条件付確率を

$$p_{ij} = P(x^{(t+1)} = j | x^{(t)} = i)$$

とする。この p_{ij} を要素とする以下のような行列 \mathbf{P} を、推移確率行列 (transition probability matrix) という。

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1J} \\ p_{21} & p_{22} & \cdots & p_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ p_{I1} & p_{I2} & \cdots & p_{IJ} \end{bmatrix} \quad (4.6)$$

(4.6) 式において、 $I = J$ であり、 \mathbf{P} は正方行列である。対角要素は、 $i = j$ の場合であり、時点 t から時点 $t+1$ となっても現在の状態に留まる場合の確率を表している。推移確率行列は、マルコフ連鎖の推移を決定する中核であり、状態空間が連続的な場合も含めて一般的に推移核 (transition kernel) とも呼ばれる。また、推移確率行列を構成する各要素 p_{ij} は確率であるため、

$$0 \leq p_{ij} \leq 1, \\ \sum_{j=1}^J p_{ij} = 1$$

という条件を満たすものとする。

確率過程の推移

時点 t において状態 i にある確率を

$$\pi_i^{(t)} = P(x^{(t)} = i) \quad (4.7)$$

とする。これをすべての状態について考え、ベクトルで表すと

$$\boldsymbol{\pi}^{(t)} = (\pi_1^{(t)}, \pi_2^{(t)}, \dots, \pi_I^{(t)})' \quad (4.8)$$

となる。(4.8) 式は時点 t における状態確率分布 (state probability distribution) と呼ばれる。 $\boldsymbol{\pi}^{(1)}$ は推移が始まる前の初期状態の分布 (初期分布) を表す。初期分布から、 $t = 2$ と時点を更新して $\boldsymbol{\pi}^{(2)}$ を求めるには、推移確率行列 \mathbf{P} を利用して

$$\boldsymbol{\pi}^{(2)} = \boldsymbol{\pi}^{(1)} \mathbf{P} \quad (4.9)$$

とすればよい。さらに、(4.9) 式と同様の手続きによって、 $\boldsymbol{\pi}^{(2)}$ を利用して $\boldsymbol{\pi}^{(3)}$ を求めることができる。

$$\boldsymbol{\pi}^{(3)} = \boldsymbol{\pi}^{(2)} \mathbf{P} \quad (4.10)$$

(4.9) 式および(4.10) 式から、時点 t と時点 $t + 1$ の状態確率分布に関して、

$$\boldsymbol{\pi}^{(t)} \mathbf{P} = \boldsymbol{\pi}^{(t+1)} \quad (4.11)$$

という関係性が導かれる。このように確率過程が更新されることを推移という。さらに、(4.10) 式の $\boldsymbol{\pi}^{(2)}$ に(4.9) 式の右辺を代入すると、

$$\boldsymbol{\pi}^{(3)} = \boldsymbol{\pi}^{(1)} \mathbf{P} \mathbf{P} \quad (4.12)$$

となることから、任意の $t + 1$ 時点における状態確率分布は、初期分布を基準として以下のように表現することもできる。

$$\boldsymbol{\pi}^{(1)} \mathbf{P}^t = \boldsymbol{\pi}^{(t+1)} \quad (4.13)$$

エルゴード性

MCMC法では、エルゴード的なマルコフ連鎖の持つ性質を利用して目標分布を得て、目的の乱数をサンプリングする。ここでは、エルゴード性を満たすための3つの条件について概説するとともに、エルゴード的なマルコフ連鎖の性質について述べる。

既約的 マルコフ連鎖の状態 i が、有限回の推移 t で状態 j へと到達する確率が0ではないとき、すなわち $P_{ij}^t > 0$ のとき、状態 i は状態 j に到達可能であるという。同時に、 $P_{ji}^t > 0$ も成り立つならば、状態 i と状態 j は互いに到達可能であるという。マルコフ連鎖の状態空間 S の要素がすべて互いに到達可能であるとき、そのマルコフ連鎖は既約的 (irreducible) であると表現される。

再帰的 状態 i から推移を開始し、再び状態 i となる最小の推移回数 (時点の数) を T_i とする。このとき、 $P(T_i < \infty) = 1$ ならば、状態 i は再帰的 (recurrent) であるという。つまり、有限回の推移回数で再び同一の状態に戻って来ることが確実であることを意味している。さらに、再帰的な状態 i に関して $E[T_i] < \infty$ であるとき、正再帰的 (positive recurrent) であるといわれる。既約的なマルコフ連鎖が正再帰的であるとき、任意の状態は推移の間に限りなく何度も到達され、同一の状態に戻るまでの時間は有限である。

非周期的 既約的なマルコフ連鎖は、状態 i から状態 j へと到達する推移の回数に規則性があるか否かによって、その性質が分類される。連鎖の各状態が同一の繰り返しになる時点間の長さを周期 (period) と呼び、すべての状態について周期が1であるマルコフ連鎖を非周期的 (aperiodic) であるという。周期的なマルコフ連鎖では、一定の周期をもって同一の連鎖の状態を繰り返すため、何度推移を繰り返そうとも初期分布の影響が残り続けることとなる。

エルゴード的なマルコフ連鎖

上述の3つの条件を満たすマルコフ連鎖、すなわち既約的、正再帰的、かつ非周期的なマルコフ連鎖は、エルゴード性 (ergodicity) を満たし、エルゴード

的であるといわれる。エルゴード的なマルコフ連鎖は、状態空間の部分集合におけるすべての状態 i, j に対して

$$\lim_{t \rightarrow \infty} P_{ij}^t = \pi(j) \quad (4.14)$$

が成立する。このことは、エルゴード的なマルコフ連鎖であれば、いかなる初期状態から推移を開始しても、推移を繰り返して、 $t \rightarrow \infty$ とすると、時点によらずある1つの分布 π に収束することを意味している。すなわち、マルコフ連鎖の推移を表す(4.11)式において

$$\pi P = \pi \quad (4.15)$$

が成り立つ。(4.15)式は、推移確率行列 P に従って状態が推移しても、全体の分布は必ず π に留まることを表している。推移を開始する時点での分布が

$$\pi^{(1)} = \pi \quad (4.16)$$

である場合、(4.11)式および(4.15)式から、

$$\pi^{(1)}, \pi^{(2)}, \dots, \pi^{(T)} = \pi \quad (4.17)$$

となり、時点 t の区別なく $\pi^{(t)} = \pi$ へと収束する。このとき、分布 π を推移確率行列 P を持つマルコフ連鎖の不変分布 (invariant distribution)、または定常分布 (stationary distribution) と呼ぶ。

エルゴード的なマルコフ連鎖では、再帰性が満たされているため、状態は対象とする集合の外へ推移していくことはない。また、既約的である故に、連鎖は不変分布に留まり続け、非周期的に状態間を推移し、部分集合のすべての要素に到達することが保障されている。つまり、エルゴード的なマルコフ連鎖では、ある期間の推移後、不変分布の収束に至ってから得られる乱数は、すべて不変分布に従っている。

モンテカルロ法

大数の法則を利用して、積分を数値的に求める方法をモンテカルロ法という(中妻, 2007)。一般に、確率変数の列 $\{x^{(t)}\} (t = 1, 2, \dots)$ と確率変数 x について、

$\epsilon > 0$ に対して

$$\lim_{t \rightarrow \infty} P(|x_t - x| \geq \epsilon) = 0 \quad (4.18)$$

または、同等に

$$\lim_{t \rightarrow \infty} P(|x_t - x| < \epsilon) = 1 \quad (4.19)$$

であるならば、 $\{x^{(t)}\}$ は x に確率収束するという。

いま、確率変数の列 $\{x^{(t)}\}$ が、相互に独立であり、それぞれ平均 μ 、分散 $\sigma^2 < \infty$ の分布に従っているとす。このとき、 $\{x^{(t)}\}$ は独立同分布に従う (independent and identically distributed; iid) という。その標本平均 $\bar{x}^{(t)} = \frac{1}{t} \sum_{i=1}^t x^{(i)}$ について、

$$\lim_{t \rightarrow \infty} P(|\bar{x}^{(t)} - \mu| < \epsilon) = 1 \quad (4.20)$$

が成立する。(4.20) 式は大数の弱法則と呼ばれる定理であり、ここから、 $t \rightarrow \infty$ のとき、標本平均は期待値 μ に確率収束し、期待値が標本平均で近似できることがわかる。

ベイズ的推測における母数の推定値として、事後分布の期待値を計算するためには、積分計算が必要とされる。積分を数値的に求める方法の1つとしてモンテカルロ法が用いられる。いま、積分の対象となる関数を $f(\mathbf{x})$ とする。もし、モンテカルロ法によって、独立に同一の分布 $f(\mathbf{x})$ に従う乱数 $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(T)}$ を生成することができれば、積分は

$$\int f(\mathbf{x}) d\mathbf{x} \approx \frac{1}{T} \{f(\mathbf{x}^{(1)}) + \dots + f(\mathbf{x}^{(T)})\} \quad (4.21)$$

のように近似することができる。大数の弱法則によって、 $T \rightarrow \infty$ において、平均による近似は対象の積分の値に確率収束することが保証される。このとき、生成された乱数 $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(T)}$ をモンテカルロ標本 (Monte Carlo sample) と呼び、それらの乱数を使って積分を数値的に求める方法をモンテカルロ積分と呼ぶ。乱数のサンプリング法としては、棄却サンプリングや重点的サンプリングなどが用いられる。

マルコフ連鎖モンテカルロ法

目標分布が高次元な場合やあまりに複雑な場合には、従来のモンテカルロ法によるサンプリングが困難となる。このようにサンプリングが困難な目標分布から乱数を発生させる際に、現在の状態に依存しながら徐々に状態空間内を推移するマルコフ連鎖が有用である。マルコフ連鎖モンテカルロ法では、収束する不変分布を目標分布としてエルゴード的なマルコフ連鎖を構成し、推移を繰り返すことで、最終的にその分布を事後分布として得ることができる。以下では、豊田 (2008) に基づき、MCMC 法の代表的なアルゴリズムとして、メトロポリス・ヘイスティングス (metropolis-hastings; MH) アルゴリズム (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953; Hastings, 1970) と、ギブスサンプラー (gibbs sampler) について説明する。

メトロポリス・ヘイスティングスアルゴリズム

MH アルゴリズムでは、詳細釣り合い条件 (detailed balance condition) と呼ばれる、マルコフ連鎖が不変分布に収束するための十分条件が重要な役割を果たす。詳細釣り合い条件は、状態空間に属する全ての x について、

$$\pi(x^{(t)})P(x^{(t+1)}|x^{(t)}) = \pi(x^{(t+1)})P(x^{(t)}|x^{(t+1)}) \quad (4.22)$$

が満たされている場合に成立する。このとき、ある状態 $x^{(t+1)}$ から状態 $x^{(t)}$ へ反転して推移する状況を全ての $x^{(t)}$ について周辺化すると 1 となる。そのため、詳細釣り合い条件は、

$$\int \pi(x^{(t)})P(x^{(t+1)}|x^{(t)})dx^{(t)} = \pi(x^{(t+1)}) \quad (4.23)$$

と同義である。(4.23) 式は時点 t から $t+1$ に推移しても、 $x^{(t+1)}$ が従う分布は $x^{(t)}$ が従っていた分布と変わらず $\pi(\cdot)$ となることを意味している。

いま、提案分布 (proposal distribution) を $q(x|x^{(t)})$ とする。 x および $x^{(t)}$ は任意なので、提案分布が詳細釣り合い条件を満たしていない状況は

$$\pi(x^{(t)})q(x|x^{(t)}) > \pi(x)q(x^{(t)}|x) \quad (4.24)$$

のように仮定できる。(4.24)式は釣り合いが崩れ、 $x^{(t)}$ は候補の x へと多く推移するが、 x から $x^{(t)}$ への推移は少ないという状況を表している。

そこで、確率 $\alpha(x|x^{(t)})$ を導入し、

$$P(x|x^{(t)}) = q(x|x^{(t)}) \times \alpha(x|x^{(t)}) \quad (4.25)$$

のように、 $x^{(t)}$ から x への推移の量を釣り合いがとれるように調整する。(4.24)式では、 x から $x^{(t)}$ への推移が少なかったため、(4.25)式の $P(x|x^{(t)})$ が詳細釣り合い条件を満たすように調整するには、 $\alpha(x^{(t)}|x)$ を1と最大にして

$$\alpha(x|x^{(t)}) = \frac{\pi(x)q(x^{(t)}|x)}{\pi(x^{(t)})q(x|x^{(t)})}$$

とする。この $\alpha(x^{(t)}|x)$ を用いて、

$$\begin{aligned} \pi(x^{(t)})q(x|x^{(t)})\alpha(x|x^{(t)}) &= \pi(x)q(x^{(t)}|x)\alpha(x^{(t)}|x) \\ &= \pi(x)q(x^{(t)}|x) \end{aligned}$$

とする。

以上のように、MHアルゴリズムは、目標分布 $\pi(x)$ からのサンプリングが困難な場合に、代わりにサンプリングの容易な提案分布から乱数を発生させ、目標分布と提案分布とのズレを詳細釣り合い条件を用いて修正するステップを加えることで、最終的に目標分布からのサンプリングを可能にする方法である。MHアルゴリズムの更更新手続きは以下である。

1. x の初期値 $x^{(0)}$ を決める。
2. $t = 1, 2, \dots$ に対して、次を繰り返す。
 - (a) 新しい候補点 x' を提案分布 $q(x|x^{(t)})$ から発生させる。
 - (b) 以下の確率を計算する。

$$\alpha(x|x^{(t)}) = \min \left[\frac{\pi(x)q(x^{(t)}|x)}{\pi(x^{(t)})q(x|x^{(t)})}, 1 \right]$$

- (c) $u \sim U(0, 1)$ を発生させ、 $x^{(t+1)}$ を以下に従って決める。

$$x^{(t+1)} = \begin{cases} x' & u \leq \alpha(x|x^{(t)}) \text{ の場合} \\ x^{(t)} & \text{それ以外の場合} \end{cases}$$

3. 十分に大きな T^* に対して, 標本 $x^{(t)} (t = T^*, \dots)$ を記録する。

MH アルゴリズムの推移核 p_{MH} は $x \in A$ のすべての集合に対して以下となる。

$$p_{MH}(A|x^{(t)}) = \int_A q(x|x^{(t)})\alpha(x|x^{(t)})dx + I(x^t \in A)r(x^{(t)})$$

$$r(x^{(t)}) = p(x^{(t)}|x^{(t)}) = 1 - \int q(x|x^{(t)})\alpha(x|x^{(t)})dx$$

MH アルゴリズムでは, 詳細釣り合い条件が満たされるように設定された推移核 p_{MH} を利用することで, 不変分布を $\pi(x)$ とするマルコフ連鎖を構成することを可能にする。

ギブスサンプラー

MH アルゴリズムの特別な場合であり, 多次元の問題を少ない次元の問題の組み合わせに帰着させることができる有用なアルゴリズムとして, ギブスサンプラー (gibbs sampler) がある。もし一部の全条件付分布が既知の分布にならず, 全条件付事後分布からのサンプリングが困難な場合は, その全条件付分布については MH アルゴリズムを適用すればよい。これはギブス内メトロポリス (metropolis within gibbs) アルゴリズム, または複合マルコフ連鎖モンテカルロ法 (hybrid MCMC method) と呼ばれる方法である。ギブスサンプラーは多重ブロック MH アルゴリズムにおいて, 各ブロックの全条件付分条件付布 (条件付提案分布) が条件付目標分布に一致しているサンプリング法であり,

$$q_j(x'_k|x_k^{(t)}, x_{-k}) = \pi(x'_k|x_{-k}) \quad (4.26)$$

$$q_k(x_k^{(t)}|x'_k, x_{-k}) = \pi(x_k^{(t)}|x_{-k}) \quad (4.27)$$

が成立する。採択確率は

$$\frac{\pi(x'_k|x_{-k})\pi(x_k^{(t)}|x_{-k})}{\pi(x_k^{(t)}|x_{-k})\pi(x'_k|x_{-k})} = 1 \quad (4.28)$$

と, 1 となり, 候補は常に採択されることとなる。

$K = 2$ の場合, ギブスサンプラーにおける推移核は

$$p_G(\mathbf{x}^{(t+1)}|\mathbf{x}^{(t)}) = \pi(x_1^{(t+1)}|x_2^{(t)})\pi(x_2^{(t+1)}|x_1^{(t+1)}) \quad (4.29)$$

であり,

$$\begin{aligned}
& \int \pi(\mathbf{x}^{(t)}) p_G(\mathbf{x}^{(t+1)} | \mathbf{x}^{(t)}) d\mathbf{x}^{(t)} \\
&= \int \int \pi(x_1^{(t)}, x_2^{(t)}) \pi(x_1^{(t+1)} | x_2^{(t)}) \pi(x_2^{(t+1)} | x_1^{(t+1)}) dx_1^{(t)} dx_2^{(t)} \\
&= \pi(x_2^{(t+1)} | x_1^{(t+1)}) \int \int \pi(x_1^{(t)}, x_2^{(t)}) \pi(x_1^{(t+1)} | x_2^{(t)}) dx_1^{(t)} dx_2^{(t)} \\
&= \pi(x_2^{(t+1)} | x_1^{(t+1)}) \int \pi(x_1^{(t)}) \pi(x_1^{(t+1)}) dx_1^{(t)} \\
&= \pi(x_2^{(t+1)} | x_1^{(t+1)}) \pi(x_1^{(t+1)}) \int \pi(x_1^{(t)}) dx_1^{(t)} \\
&= \pi(x_2^{(t+1)} | x_1^{(t+1)}) \pi(x_1^{(t+1)}) \\
&= \pi(x_1^{(t+1)}, x_2^{(t+1)}) \tag{4.30}
\end{aligned}$$

となる。これはギブスサンプラーによって不変分布に従って推移が行われることを表している。

4.3 ハミルトニアンモンテカルロ法

本節では、HMC法について概説する。次章で論じる研究IIIでは、このHMC法による推定が行われる。ハミルトニアンモンテカルロ法はサンプリングアルゴリズムにおける設定が一般的なMCMC法よりも比較的難しいものの、適切に設定を行うことができれば、従来のMCMC法よりも収束が速まるという特徴を有している。さらに、例えば前節で述べたギブスサンプリングによる方法を適用するには、他のすべての母数が与えられたときの当該母数に関する事後分布を提示する必要があったが、HMC法では、対数事後分布の母数に関する微分が必要なだけであり、その過程において煩雑な計算を必要としないことも利点の1つである。

ハミルトン関数

HMC法は、分子動力学法におけるハミルトン力学 (Hamiltonian dynamics) と、モンテカルロ法におけるメトロポリス法を融合した手法であるため、ハイブリッ

ドモンテカルロ (hybrid Monte Carlo) 法とも呼ばれる。ここではまず、ハミルトニアン力学の根底を成すハミルトン関数について概説する。

力学の観点から、仮想的な物体の位置変数ベクトル $\boldsymbol{x} = (x_1, \dots, x_n)$ の値は時間 t の経過に伴い変化すると考え、位置変数ベクトルの各要素に対応する運動量変数ベクトルを $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$ とする。位置変数と運動量変数による結合空間は位相空間と呼ばれる。

物体の運動に関して、位置エネルギー $U(\boldsymbol{x})$ は位置変数ベクトル \boldsymbol{x} によって、運動エネルギー $K(\boldsymbol{\delta})$ は運動量変数ベクトル $\boldsymbol{\delta}$ によって与えられる。エネルギー保存則により、系 (system) の全エネルギーは位置エネルギーと運動エネルギーの和として与えられる。

$$H(\boldsymbol{x}, \boldsymbol{\delta}) = U(\boldsymbol{x}) + K(\boldsymbol{\delta}) \quad (4.31)$$

この (4.31) 式をハミルトン関数 (Hamiltonian function; Neal, 2011), もしくはハミルトニアン (松本, 2004) と呼ぶ。ハミルトン関数を利用した1次連立微分方程式

$$\frac{dx_i}{dt} = \frac{\partial H}{\partial \delta_i}, \quad (4.32)$$

$$\frac{d\delta_i}{dt} = -\frac{\partial H}{\partial x_i} \quad (4.33)$$

はハミルトンの運動方程式と呼ばれる。(4.32)式, (4.33)式により与えられた方程式を解くことで、物体の運動の変化を知ることが可能となる。

ハミルトン力学には、時間反転性 (reversibility), ハミルトニアン H の不変性 (conservation of the Hamiltonian), 位相空間の体積保存 (volume preservation) という3つの重要な性質がある。以下では、これらの性質の1つ1つを簡単に説明する。

時間反転性 時間反転性とは、ある時点 t から d だけ経過した時点 $t+d$ において、速度 $\boldsymbol{\delta}$ を反転させ、 $-\boldsymbol{\delta}$ とすると、それまで通った軌跡とまったく同じ軌跡を逆にたどって、元の時点 t に戻ってくる性質である。時点 t における状態 $(\boldsymbol{x}(t), \boldsymbol{\delta}(t))$ と、時点 $t+s$ における状態 $(\boldsymbol{x}(t+s), \boldsymbol{\delta}(t+s))$ は一対一変換、すなわち写像の関係にあり、その逆写像が必ず存在する。また、写像 $g : (\boldsymbol{x}', \boldsymbol{\delta}') = g(\boldsymbol{x}, \boldsymbol{\delta})$ に関して、 $(\boldsymbol{x}, -\boldsymbol{\delta}) = g(\boldsymbol{x}', \boldsymbol{\delta}')$ が成立する。

ハミルトニアン H の不変性 H の不変性はエネルギー保存則から直ちに導かれる。ハミルトン力学によって提案される候補点をメトロポリス法によって採択するか否かを決定する際に、 H が不変であれば採択確率は必ず1となる。ただし、実際には数値積分を行うため、必ずしも H が不変であるとは限らない。

位相空間の体積保存 体積保存は写像 g のヤコビ行列の行列式が1になることと等しい (Neal, 2011)。この性質は、位相空間上の同時分布 $p(\mathbf{x}, \boldsymbol{\delta})$ が時間経過の下で不変であることを意味している。

リープフロッグ法

ハミルトニアン H の不変性において述べたように、(4.32)式および(4.33)式で与えられたハミルトンの運動方程式を解くためには、数値積分を利用して近似的にその解を求めることとなる。このとき、前述した3つの性質を保ったままマルコフ連鎖の更新を行わなくてはならない。ここでは、体積保存の性質を損なわずに数値積分を行う方法として、リープフロッグ (leap-frog) 法について述べる。

リープフロッグ法では、まず運動量変数をステップサイズ $\epsilon/2$ として半ステップだけ更新し、次にステップサイズ ϵ による位置変数の更新を行う。初期値 $\mathbf{x}^{(0)}, \mathbf{p}^{(0)}$ を定め、具体的には以下の更新式に従う。

$$\delta_i(t + \epsilon/2) = \delta_i(t) - \frac{\epsilon}{2} \frac{\partial U}{\partial x_i}[x(t)], \quad (4.34)$$

$$x_i(t + \epsilon) = x_i(t) + \epsilon \frac{\delta_i(t + \epsilon/2)}{m_i} \quad (4.35)$$

$$\delta_i(t + \epsilon) = \delta_i(t + \epsilon/2) - \frac{\epsilon}{2} \frac{\partial U}{\partial x_i}[x(t + \epsilon)] \quad (4.36)$$

これを $L = T/\epsilon$ 回繰り返す。時点 T における位置変数ベクトルと運動量変数ベクトル $\mathbf{x}^{(T)}, \mathbf{p}^{(T)}$ を得る。

ハミルトン力学を利用したMCMC法

位置変数 \mathbf{x} と運動量変数 $\boldsymbol{\delta}$ の位相空間における同時分布を

$$p(\mathbf{x}, \boldsymbol{\delta}) = \frac{1}{Z_H} \exp[-H(\mathbf{x}, \boldsymbol{\delta})/\iota] \quad (4.37)$$

$$= \frac{1}{Z_H} \exp[-U(\mathbf{x})/\iota] \exp[-K(\boldsymbol{\delta})/\iota] \quad (4.38)$$

と定義する。

$$p(\mathbf{z}) = \frac{1}{Z} \exp[-E(\mathbf{x})/\iota] \quad (4.39)$$

による表現はカノニカル分布 (canonical distribution) と呼ばれる。ここで $E(\mathbf{x})$ はエネルギーを表す。 Z は正規化定数¹、 ι は系の温度 (temperature of the system; Neal, 2011, p.123) と呼ばれ、ここでは $\iota = 1$ とする。

MCMC の観点からは、位置変数 \mathbf{x} がパラメタであり、その事後分布が目標分布となる。いま事前分布 $\pi(\mathbf{x})$ と尤度関数 $L(\mathbf{D}|\mathbf{x})$ により、パラメタ \mathbf{x} の事後分布が $\pi(\mathbf{x}|\mathbf{D}) \propto L(\mathbf{D}|\mathbf{x})\pi(\mathbf{x})$ と表されるとすると、位置エネルギーの事後分布を

$$U(\mathbf{x}) = -\log[L(\mathbf{D}|\mathbf{x})\pi(\mathbf{x})] \quad (4.40)$$

のように定義できる。また、運動エネルギーは一般に

$$K(\boldsymbol{\delta}) = \sum_{i=1}^n \frac{\delta_i^2}{2m_i} \quad (4.41)$$

と定義される。ここで m は質量に対応するパラメタである (松本, 2004, p.149, Neal, 2011, p.114)。(4.40) 式と (4.41) 式を用いて同時分布 (4.38) 式から乱数を発生させることで、HMC 法を用いた MCMC 標本を得ることが可能となる。

ハミルトニアンモンテカルロ法のアルゴリズム

HMC 法のアルゴリズムを以下に示す。

1. \mathbf{x} の初期値 $\mathbf{x}^{(0)}$ 、ステップサイズ ϵ 、および更新回数 L を定める。
2. 時点 $t = 1, 2, \dots, T$ において、以下の手順を繰り返す。

¹ $p(\mathbf{z})$ の和もしくは積分を 1 に調整するために導入される (Neal, 2011, p.123)。

- (a) 運動量変数に関する初期値 $\boldsymbol{\delta}^{(0)}$ を多変量正規分布 $N(\mathbf{0}, \mathbf{I}_n)$ より発生させる。
- (b) $\mathbf{x}^{(t-1)}$ と $\boldsymbol{\delta}^{(0)}$ を初期値とし、リープフロッグ法により、候補点 $\tilde{\mathbf{x}}, \tilde{\boldsymbol{\delta}}$ を求める。
- (c) 採択確率 α を計算する。

$$\alpha = \min\{1, \exp[-H(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\delta}}) + H(\mathbf{x}^{(t-1)}, \boldsymbol{\delta}^{(0)})]\} \quad (4.42)$$

- (d) $u \sim U(0, 1)$ を発生させ、 $\mathbf{x}^{(t)}$ を以下に従って決める。

$$\mathbf{x}^{(t)} = \begin{cases} \tilde{\mathbf{x}} & u \leq \alpha \text{ の場合} \\ \mathbf{x}^{(t-1)} & \text{それ以外の場合} \end{cases} \quad (4.43)$$

ハミルトニアン H の値は常に一定であるため、ハミルトン力学の仮定の下では時間経過によって $p(\mathbf{x}, \boldsymbol{\delta})$ からエルゴード的なマルコフ連鎖を構成することができない。これに対して、アルゴリズム 2(a) は、分布 $p(\mathbf{x}, \boldsymbol{\delta})$ を不変に保ちつつ、 H の値を変えるような位相空間での追加的な移動の導入を目的とするものであり、エルゴード的なマルコフ連鎖構成のために必要な手順である。また、アルゴリズム 2(c) は、 H の不変性が、数値積分を行ったことによる誤差が原因で満たされないことを防ぐための手順である。この更新手順において、誤差によるバイアスを取り除くことで、不変分布からのサンプリングが保たれる。このステップによって、詳細釣り合い条件が満たされることが保証される²。

なお、事後統計量を算出する際には、同時分布 $p(\mathbf{x}, \boldsymbol{\delta})$ から乱数を発生させ、運動量変数に関する乱数を破棄し、位置変数に関するサンプルのみを利用する。

4.4 収束判定

MCMC 法の利用においては、他の伝統的な統計手法を適用する場合と同様に、得られた推定値について推定が問題なく成功したのかどうか、分析の妥当性を検討する手続きまで含めて、分析結果の解釈を行う必要がある。

MCMC 法による推測を行う場合には、生成したマルコフ連鎖のうち、確かに不変分布からのサンプリングができていると考えられる MCMC 標本のみを推

²詳細釣り合い条件が成立することの証明は Neal (2011) において確認することができる。

定の材料とする必要がある。しかし、マルコフ連鎖の状態は、系列における1つ前の状態に依存して決定されるため、不変分布からのサンプリングではない値から生成を開始したマルコフ連鎖では、しばらくの間その初期値の影響が残り、構成要素が不変分布からの無作為標本とみなせる状態になるためには推移が複数回繰り返されることが必要となる。このように初期値の影響がなくなり、マルコフ連鎖が不変分布に至ることを連鎖の収束と呼ぶ。

マルコフ連鎖が収束して以降のMCMC標本のみを利用して推測を行うために、十分な更新回数を経た後の連鎖から、初期の構成要素を取り除いた上で、不変分布からの無作為標本と見なせる残りの要素を用いて事後統計量の算出などが行われる。この破棄される更新期間は、バーンイン期間 (burn-in period)、あるいはウォームアップ期間 (warm-up period) と呼ばれる。実際の利用場面では、バーンイン期間の長さおよび更新回数の上限は分析者が任意に決定することとなり、連鎖の発生に先立って適切なバーンイン期間の長さを決定するための方法や基準等は未だ示されていない。一方で、任意のバーンイン期間を除いた残りの要素が収束しているか否かを判定するための方法については、様々な指標が提案されている。

時系列プロット・自己相関関数図

収束を判定するためのもっとも簡単かつ有効な方法は、マルコフ連鎖を構成する要素の値そのものを時系列順に図示することである。すなわち、横軸にサンプリングの繰り返し回数、縦軸に連鎖要素の値を配したグラフを描き、観察すればよい。このようなグラフは、連鎖要素の時系列プロット、あるいはトレースプロットと呼ばれる。不変分布に収束したマルコフ連鎖の状態空間は、推移を何回重ねても同じ状態に留まる。したがって、時系列プロットにおいて値の変動に系列的な偏りが見られず、不変分布の最頻値付近を行ったり来たりするような状態になっていれば、概ね収束したと見なして問題ない。

バーンイン期間も含めて時系列プロットを図示することで、初期段階では初期値の影響によって偏っている軌跡が、推移を重ねるに連れて一定の値の付近に収束していく様子が確認できる。そこで、時系列プロットにおいて値が安定

した以降の要素のみをパラメタの推定に用いるようにバーンイン期間を設定することで、事後的な調整を行うことが可能となる。逆に、任意に決定したバーンイン期間以降の要素のみを用いて時系列プロットを描き、その系列に系統的な偏りが存在していないかどうか黙視確認することで、事前に定めたバーンイン期間が適切であったかどうかを検討することもできる。

また、時系列プロットの他に、連鎖における1つ前の時点の値との自己相関係数を計算し図示する自己相関関数図も、収束判定のために利用されることが多い。マルコフ連鎖の値は直前の要素に依存して決定されるため、近い間隔の要素との自己相関係数が0になることはない。しかし一方で、MCMC法において必要とされるマルコフ連鎖は、目標分布からの無作為標本を再現するような要素によって構成されるものであり、ある程度以上間隔が離れている要素同士は互いに独立であることが期待される。そこで、自己相関関数図を描き、その値が比較的近い間隔において0に収束しているかどうかを確認することで、マルコフ連鎖の収束判定に役立てることができる。

収束判定指標 \hat{R}

以下では、連鎖の収束を事後的に確認するための1つの指標として、Gelman & Rubin (1992) による \hat{R} の算出方法について述べる。なお、表記については Gelman & Rubin (1992), Gelman (1996) および豊田 (2008) に従った。Gelman & Rubin (1992) は、複数の連鎖を発生させ、それらの分散を比較に用いることで連鎖の収束を判定する方法およびその指標 \hat{R} を提案した。 \hat{R} は estimated potential scale reduction (EPSR; Gelman & Rubin, 1992) とも呼ばれる。

いま、同じパラメタに対して任意の異なる初期値から発生させた K 本の連鎖を構成する状況を考える。パラメタ θ に関する、事後分布からのサンプリング標本を用いて、更新期間 T の連鎖を

$$\{\theta_k^{(1)}, \theta_k^{(2)}, \dots, \theta_k^{(T)}\}, \quad k = 1, \dots, K \quad (4.44)$$

のように K 本構成する。これらの連鎖から、共通に最初の m 回分をバーンイン期間として破棄したとき、連鎖の残りが収束していると言えるか否かを判定

することを目的とする。以降ではバーンイン期間破棄後の連鎖の要素を全要素として扱う。

まず、異なる連鎖間における θ の値のばらつき $\text{var}_B(\theta)$ および、ある 1 本のマルコフ連鎖 (k 番目のマルコフ連鎖) 内における値のばらつきの推定値 $\text{var}_W(\theta)$ を以下の式に従って計算する。

$$\widehat{\text{var}}_B(\theta) = \frac{T-m}{K-1} \sum_{k=1}^K (\bar{\theta}_k^{(\cdot)} - \bar{\theta}^{(\cdot)})^2 \quad (4.45)$$

$$\widehat{\text{var}}_W(\theta) = \frac{1}{K(T-m-1)} \sum_{k=1}^K \sum_{t=1}^{T-m} (\theta_k^{(t)} - \bar{\theta}^{(\cdot)})^2 \quad (4.46)$$

(4.45) 式は系列内平均 (with-in sequence mean), (4.46) 式は系列内分散 (within-sequence variance) と呼ばれる。ここで、式中の $\bar{\theta}_k^{(\cdot)}$ は k 番目の連鎖に含まれる m 個を破棄した後の全要素の平均

$$\bar{\theta}_k^{(\cdot)} = \frac{1}{T-m} \sum_{t=1}^{T-m} \theta_{kt} \quad (4.47)$$

を表している。また、 $\bar{\theta}^{(\cdot)}$ は K 本の連鎖ごとの K 個の全要素平均であり、

$$\bar{\theta}^{(\cdot)} = \frac{1}{K} \sum_{k=1}^K \bar{\theta}_k^{(\cdot)} \quad (4.48)$$

と表される。

これら分散成分 $\widehat{\text{var}}_B(\theta)$, $\widehat{\text{var}}_W(\theta)$ を用いて、目標分布のパラメタ θ の真の分散 $\text{var}(\theta)$ を推定する。1 つ目は、以下を真の分散の推定値とする方法である。

$$\widehat{\text{var}}(\theta) = \frac{T-m-1}{T-m} \widehat{\text{var}}_W(\theta) + \frac{1}{T-m} \widehat{\text{var}}_B(\theta) \quad (4.49)$$

(4.49) 式は各連鎖が目標分布に収束している場合には不偏推定量となる。Gelman & Rubin (1996) は $\text{var}(\theta)$ を過剰な散らばりに対する保守的な推定量と呼んでいる。 K 本の連鎖のうち非収束の連鎖が 1 本でもある場合には、連鎖間で同一要素の値が異なり得るため、 $\widehat{\text{var}}_B(\theta)$ は大きな値となる。結果として、マルコフ連鎖が非収束であるならば、(4.49) 式の値は真の $\text{var}(\theta)$ より大きくなってしまふことが考えられる。

これに対して、真の分散 $\text{var}(\theta)$ を推定するための3つ目の方法として、 $\widehat{\text{var}}(\theta)$ と $\widehat{\text{var}}_W(\theta)$ の比

$$\hat{R} = \sqrt{\frac{\widehat{\text{var}}(\theta)}{\widehat{\text{var}}_W(\theta)}} \quad (4.50)$$

を指標として用いる方法が提案された。推測に用いた連鎖が全て同一の不変分布に収束している場合には、 $\widehat{\text{var}}_W(\theta)$ の値が真の分散 $\text{var}(\theta)$ に等しくなることが予想される。つまり、すべての連鎖が完全に収束しているならば、(4.50)式を構成する要素はどちらも $\text{var}(\theta)$ に一致し、このとき $\hat{R} = 1$ となる。一方で、連鎖の更新回数が不十分であり、推測に用いる連鎖の内に非収束の連鎖が含まれている場合には、目標分布の全範囲からのサンプリングが十分に行われていないことが予想されるため、 $\widehat{\text{var}}_W(\theta)$ の値は真値 $\text{var}(\theta)$ に対して過少に推定されることとなる。したがって、 \hat{R} が十分に1に近ければマルコフ連鎖は収束しているものと判断し、そうでないならば連鎖は非収束であると判断可能である。Gelman (1996) では、シミュレーションを通じて、 \hat{R} が1.2以上の場合には、非収束が示唆される結果を得たことから、 \hat{R} の値が1.2、もしくは1.1よりも小さければ連鎖は収束したものとする判断基準を提案している。

なお、構成した連鎖が1本の場合は、要素を複数に分割し、それらの要素集合を擬似的に構成された複数連鎖と見なし上述の手順を適用することで、収束判定を行うことが可能である。

第5章 研究III: 同一立場内に複数の評価者がいる多特性多評価者データにおける信頼性と妥当性の検討

第5章では、第3章の研究IIから引き続き、評価者の違いを“方法”とするMTMRデータのための分析方法について論じる。本章では最初に、MTMRデータを用いて測定の信頼性や妥当性を検討する際に特有の問題について詳しく述べる。具体的には、同一立場内に複数の評価者が存在する場合に、従来の分析手法では評価者の人数が信頼性や妥当性に与える影響を適切に考慮できていないという点について指摘する。この問題に対処するため、研究IIIでは、同一立場内に複数の評価者が存在するMTMRデータを分析するための新たな方法を提案する。この新たなアプローチは、HMC法を導入したマルコフ連鎖モンテカルロ法を利用した方法である。本章の最後では、研究IIでも用いた実データを分析対象として適用例を示し、提案手法の利点および実用可能性について論じる。

5.1 問題と目的

MTMRデータにおける、測定方法や測定機会を“方法”として扱うMTMMデータとの大きな違いは、同一“方法”内で複数の測定値が得られる可能性があるという点である。例えば360度フィードバックでは、1人の被評価者につき、複数の同僚、あるいは複数の部下から評定結果が得られる場合がある。すべての被評価者について同じ人数の他者評価が得られているMTMRデータに対して、確認的因子分析モデルを適用した例として、Mount et al. (1998)が挙げられる。Mount et al. (1998)では、すべての被評価者について自己評価に加え、それ

ぞれ2人の上司，同僚，部下から評価が得られており，7つの評価者因子（自己評価，上司評価1，上司評価2，同僚評価1，同僚評価2，部下評価1，部下評価2）を仮定したモデルを当てはめている。しかしながら，選ばれる2人の上司，同僚，部下はいずれも被評価者によって異なるはずであり，交換可能な2人の他者からの評価について，単純に人数ぶんの評価者因子を仮定することが適切であるとは言い難い。

さらに，実際の360度フィードバックでは，得られる同僚評価の数，あるいは部下評価の数は，被評価者によって異なることが多い。このように複数人から他者評価が得られたMTMRデータに対して，従来のMTMMデータのための確認的因子分析モデルを適用する際には，慣例的に平均値を観測変数として用いてMTMM行列が計算されてきた(Forsythe, McGaghie, & Friendman, 1986; Lawler, 1967; Mount, 1984)。本来，評価者の人数は測定の信頼性に直接寄与する要因であると考えられる。しかし，単純に平均値を計算し，それをそのまま分析に用いるという従来の方法では，労力を費やし収集した評価者の人数に関する情報を，適切に分析結果に反映することができていない。

本研究では，まず，同一立場内に複数の評価者が存在する場合に，その平均値を用いて確認的因子分析モデルを実行するとき，評価者の人数に応じて分散成分がどのように変化するかを数学的に導出した。その結果，評価者の人数の影響は，誤差分散の値の違いとして現れることが明らかとなった。平均値を利用した相関行列に従来のMTMM行列のための確認的因子分析モデルをそのまま適用すると，結果として，誤差分散が過小評価される。誤差分散の過小評価は，測定の信頼性と妥当性に関する評価に直接的に影響する。なぜなら，MTMM行列に対して確認的因子分析モデルを適用し，信頼性および収束的妥当性，弁別的妥当性について検討する際には，分散成分の分解により信頼性と妥当性が定義されるからである。そのため，誤差分散が正しく推定できていない場合には，測定の信頼性と妥当性の解釈も誤ったものになることが懸念される。

以上より，研究IIIの目的は，MTMRデータに対する確認的因子分析モデルの適用において，評価者の人数を適切に評価し，測定の信頼性と妥当性を検討するための方法を提案することとする。評価者の人数を統計的に処理するという目的のために，まず一般的に考え得る方法として，多母集団同時解析モデルの

利用が挙げられる。本研究の目的を達成するためには、評価者の人数の違いに応じてあらかじめデータを群分けし、誤差分散のみ群ごとに値が異なるモデルを構成し、それ以外の母数については群間で等値の制約を置いて推定を実行すればよい。しかしながら、多母集団同時解析モデルを安定的に推定するためには比較的大きな標本サイズを必要とし、特定の群のみ標本サイズが極端に小さい場合には信頼できる推定結果を得ることができないという問題がある。さらに、例えば同僚評価と部下評価の両方で、複数人からの評価が得られているといった状況では、同僚人数と部下人数の組み合わせで群分けをする必要があるため、群の数が積の勢いで増え、一方で各群の標本サイズは小さくなる。そこで、小さな標本サイズでも、また特定の人数の他者評価を得ている被評価者が仮にたった1人しかいない場合でも分析可能な方法として、本研究ではMCMC法によるアプローチを提案する。

適用例では、研究IIに引き続き、産業能率大学総合研究所による“ビジネス基礎力診断 S-BASE”の実施により得られた360度フィードバックの実データを用いた。“ビジネス基礎力診断 S-BASE”では、自己、上司、同僚という3つの立場からの評価を行っており、このデータでは、同僚評価について、各被評価者につき最大6人からの評定結果が得られていた。MCMC法によるアプローチを適用した結果を、従来の確認的因子分析モデルと、多母集団解析モデルによるアプローチの適用結果と比較することで、被評価者ごとに異なる同僚評価の人数に対して、統計的に適切な処理が実現できていることを確認した。

5.2 モデル

研究IIIにおいても、研究IIと同様に、CT-UMモデルを当てはめる。

観測変数として平均値を用いた場合の分散

ここで、特定の評価者の立場において、同一立場内に複数の評価者が存在する状況を考える。さらに、評価者の人数は被評価者によって異なるものとする。

n 番目のオブザーベーションについて、 $i(=1, \dots, t)$ 番目の特性の $j(=1, \dots, m)$ 番目の評価者の立場による測定値が $A^{(n)}(=1, \dots, A)$ 人の評価者から得られて

いるとする。Aは手元のデータにおける、同一立場内の評価者の人数の最大値である。A⁽ⁿ⁾は、n(= 1, ..., N)が決まれば一意に定まり、評価者の数はオブザベーションによって異なることを表している。

任意のnに注目するとき、

$$\mathbf{x} = \mathbf{\Lambda}_T \mathbf{f}_T + \mathbf{\Lambda}_M \mathbf{f}_M + \mathbf{e} \quad (5.1)$$

というモデルにしたがって生成される1つ1つの観測変数は、タウ等価測定の仮定の下で以下のように表される(a = 1, ..., A⁽ⁿ⁾)。

$$x_{ija} = \lambda_{T_{ij}} f_{T_i} + \lambda_{M_{ij}} f_{M_j} + e_{ija} \quad (5.2)$$

したがって、このA⁽ⁿ⁾人からの評価の和は以下となる。

$$\begin{aligned} x_{ij} &= \sum_{a=1}^{A^{(n)}} x_{ija} \\ &= \sum_{a=1}^{A^{(n)}} (\lambda_{T_{ij}} f_{T_i} + \lambda_{M_{ij}} f_{M_j} + e_{ija}) \\ &= A^{(n)} \lambda_{T_{ij}} f_{T_i} + A^{(n)} \lambda_{M_{ij}} f_{M_j} + \sum_{a=1}^{A^{(n)}} e_{ija} \end{aligned} \quad (5.3)$$

次に、(5.3)式の和得点から、平均点を計算する。

$$\begin{aligned} x_{ij}^* &= \frac{1}{A^{(n)}} x_{ij} \\ &= \frac{1}{A^{(n)}} (A^{(n)} \lambda_{T_{ij}} f_{T_i} + A^{(n)} \lambda_{M_{ij}} f_{M_j} + \sum_{a=1}^{A^{(n)}} e_{ija}) \\ &= \lambda_{T_{ij}} f_{T_i} + \lambda_{M_{ij}} f_{M_j} + \frac{1}{A^{(n)}} \sum_{a=1}^{A^{(n)}} e_{ija} \end{aligned} \quad (5.4)$$

(5.4)式より、もし、観測変数として平均点を利用した場合には、因子負荷量は変わらず、誤差項がA⁽ⁿ⁾人分の和の1/A⁽ⁿ⁾倍となることがわかる。

ここで、x_{ij}^{*}の分散を考える。

$$V[x_{ij}^*] = \lambda_{T_{ij}}^2 V[f_{T_i}] + \lambda_{M_{ij}}^2 V[f_{M_j}] + \left(\frac{1}{A^{(n)}} \right)^2 V \left[\sum_{a=1}^{A^{(n)}} e_{ija} \right] \quad (5.5)$$

$V[X] = E[X^2] - (E[X])^2$ より、最終項について、 $V[\sum_{a=1}^{A^{(n)}} e_{ija}] = E[(\sum e_{ija})^2] - (E[\sum e_{ija}])^2$ となる。多変量モデルの一般的な仮定である $E[e] = 0$ より、 $(E[\sum e_{ija}])^2$ は 0 となる。したがって、

$$\begin{aligned} V \left[\sum_{a=1}^{A^{(n)}} e_{ija} \right] &= E \left[\left(\sum_{a=1}^{A^{(n)}} e_{ija} \right)^2 \right] = E \left[\sum_{a=1}^{A^{(n)}} \sum_{a'=1}^{A^{(n)}} e_{ija} \times e_{ija'} \right] \\ &= E[e_{ij1}^2] + E[e_{ij2}^2] + \cdots + E[e_{ijA^{(n)}}^2] \end{aligned}$$

と表現することができる。ここで、同じ立場内の任意の 2 人の評価者の測定誤差は無相関であるという仮定を導入すると、 $a \neq a'$ のとき $E[e_{ija}e_{ija'}] = 0$ となるので、 $a = a'$ の項のみ残る。 $V[e] = E[(e - E[e])^2] = E[e^2] - 2eE[e] + (E[e])^2 = E[e^2]$ を利用すると、

$$E[e_{ij1}^2] + E[e_{ij2}^2] + \cdots + E[e_{ijA^{(n)}}^2] = V[e_{ij1}] + V[e_{ij2}] + \cdots + V[e_{ijA^{(n)}}]$$

であり、 $V[e_{ij}] = V[e_{ij1}] = \cdots = V[e_{ijA^{(n)}}]$ とすると、 $V[\sum_{a=1}^{A^{(n)}} e_{ija}] = A^{(n)} \times V[e_{ij}]$ となる。

この結果を (5.5) 式に代入すると、以下となる。

$$V[x_{ij}^*] = \lambda_{T_{ij}}^2 V[f_{T_i}] + \lambda_{M_{ij}}^2 V[f_{M_j}] + \frac{1}{A^{(n)}} V[e_{ij}] \quad (5.6)$$

したがって、同一立場内に評価者が複数いる場合、その平均値における誤差分散は $1/A^{(n)}$ 倍となることが導かれた。

マルコフ連鎖モンテカルロ法によるアプローチ

同一立場内の評価者の人数に応じて被評価者ごとに異なる誤差分散を CT-UM モデルに反映させるためのアプローチとして、MCMC 法を利用したモデルの推定を提案する。なお、以下では、自己・上司・同僚の 3 つの立場からの評価を行い、そのうち同僚評価のみ複数の回答が得られている状況を考える。

各パラメータの事前分布は以下のように設定した。

$$\begin{aligned} x_{nij} &\sim \text{Multi-Normal}(\mathbf{0}, \mathbf{\Lambda}_T \mathbf{\Phi}_T \mathbf{\Lambda}'_T + \mathbf{\Lambda}_M \mathbf{\Lambda}'_M + \mathbf{\Psi}_n) \quad (5.7) \\ \sigma_e^2 &\sim \text{Inv-Gamma}(0.001, 0.001), \quad \mathbf{\Phi}_T \sim \mathbf{U}(-1, 1) \\ \lambda_{T_{ij}} &\sim N(0, 10^6), \quad \lambda_{M_{ij}} \sim N(0, 10^6) \end{aligned}$$

ここで、 x_{nij} は n 番目のオブザベーションにおける、 $i(= 1, \dots, t)$ 番目の特性の $j(= 1, \dots, m)$ 番目の評価者の立場による測定値を表しており、同一立場内に複数の評価者がいる場合には、その平均値を測定値とする。(5.7) 式の Ψ_n は、誤差分散がオブザベーションごとに異なることを表しており、同僚評価に関する誤差分散に関して σ_e^2 を $A^{(n)}$ で除したものを Ψ_n の対角要素として用いた。

本研究では HMC 法を実装した Stan によって母数の推定を行った。EM アルゴリズムによる方法を適用するには、暫定的に推定されている母数の分布のもとでの対数尤度関数の条件付き期待値を提示する必要がある (Dempster, Laird & Rubin, 1977)。また、ギブスサンプリングによる方法を適用するには、他のすべての母数が与えられたときの当該母数に関する事後分布を提示する必要がある。それに対して、HMC 法は、対数事後分布の母数に関する微分が必要なだけである。しかもその微分は Stan が自動的に計算するので利用者は特定する必要がない。モデルを特定するだけで母数が推定できる点が HMC 法の優れた性質である。

多母集団同時解析によるアプローチ

(5.6) 式で導出した、同一立場内の評価者の人数に応じて異なる誤差分散を CT-UM モデルに反映させるためのもう 1 つの方法として、多母集団同時解析によるアプローチが考えられる。

誤差分散の導出で見たように、得られた同僚評価の人数 $A^{(n)}$ の最大値は A である。多母集団同時解析モデルによるアプローチでは、まず、 $A^{(n)}$ に応じて、オブザベーション（被評価者）を A 群のグループに分ける。すなわち、第 1 群は、1 人の同僚評価のみが得られているグループ、第 2 群は 2 人からの同僚評価が得られているグループ、 \dots 、第 A 群は A 人からの同僚評価が得られているグループである。これら A 群のデータに対して、多母集団同時解析を実行する。ただし、(5.6) 式より、モデルには以下のような制約を課す。

- 同僚評価のみ、その人数に応じて誤差分散を $1/A^{(n)}$ 倍し、集団間で異なるという制約を置く。
- 同僚以外の誤差係数はすべての群間で等値とする。

- 切片および因子負荷量はすべての群間で等値とする。
- 因子の分散および因子間相関は群間で等値とする。

信頼性と妥当性の解釈

提案手法による推定値を利用して、研究 II と同様に信頼性係数、収束的妥当性係数、非弁別的妥当性係数を計算することで、信頼性と妥当性について解釈を行う。係数の算出方法について以下に再掲した。

$$\text{Rel}(x_{ij}) = \frac{\lambda_{T_{ij}}^2 \text{var}(f_{T_i}) + \lambda_{M_{ij}}^2 \text{var}(f_{M_j})}{\lambda_{T_{ij}}^2 \text{var}(f_{T_i}) + \lambda_{M_{ij}}^2 \text{var}(f_{M_j}) + \text{var}(e_{ij})} \quad (5.8)$$

$$\text{Con}(x_{ij}) = \frac{\lambda_{T_{ij}}^2 \text{var}(f_{T_i})}{\lambda_{T_{ij}}^2 \text{var}(f_{T_i}) + \lambda_{M_{ij}}^2 \text{var}(f_{M_j}) + \text{var}(e_{ij})} \quad (5.9)$$

$$\text{Dis}(x_{ij}) = \frac{\lambda_{M_{ij}}^2 \text{var}(f_{M_j})}{\lambda_{T_{ij}}^2 \text{var}(f_{T_i}) + \lambda_{M_{ij}}^2 \text{var}(f_{M_j}) + \text{var}(e_{ij})} \quad (5.10)$$

5.3 適用例

産業能率大学の人材アセスメントツールの1つである、“ビジネス基礎力診断 S-BASE”による360度フィードバックの結果を用いて適用例を示す。“ビジネス基礎力診断 S-BASE”の実施概要および、データから計算される相関行列については、第3章で示した通りである。ただし、研究 III においては、変数ごとに平均による中心化を施したデータを用いた。

ビジネス基礎力診断データでは、“同僚評価”は1人の被評価者につき、1人～6人からの評定が得られていた。ここで、“同僚評価”の数とそれぞれのカテゴリにおけるオブザベーション数について表5.1にまとめた。2人の同僚から評価されている被評価者がもっとも多く、全体の約42%を占めた。一方で、6人の同僚からの評価を得た被評価者は1人だけであった。

表 5.1 同僚評価の数とオブザベーション数

同僚人数	1	2	3	4	5	6	計
オブザベーション数	29	169	59	47	96	1	401

結果と考察

最初に、通常のCT-UMモデルによる推定結果を表5.2に示した¹。ただし、推定においては、モデル識別のため、すべての因子の分散を1に固定した²。適合度指標の値は、CFI=0.994, TLI=0.985, RMSEA=0.063であった。非標準化解では、“同僚評価”に関する誤差分散 $\text{var}(e_{13})$, $\text{var}(e_{23})$, $\text{var}(e_{33})$ の値が、他の変数の誤差分散に比較して顕著に小さいことが確認できる。

表 5.2 CT-UM モデルによる推定結果（非標準化解）

母数	非標準化解	SE	z 値	母数	非標準化解	SE	z 値
$\lambda_{T_{11}}$	6.904	0.693	9.957	$\text{var}(e_{11})$	34.419	4.305	7.995
$\lambda_{T_{21}}$	6.364	0.614	10.357	$\text{var}(e_{21})$	22.831	3.517	6.492
$\lambda_{T_{31}}$	6.156	0.619	9.951	$\text{var}(e_{31})$	36.800	3.802	9.680
$\lambda_{T_{12}}$	9.098	0.832	10.941	$\text{var}(e_{12})$	12.594	2.605	4.834
$\lambda_{T_{22}}$	9.679	0.868	11.152	$\text{var}(e_{22})$	21.267	3.148	6.756
$\lambda_{T_{32}}$	8.930	0.776	11.506	$\text{var}(e_{32})$	18.542	3.132	5.920
$\lambda_{T_{13}}$	7.705	0.723	10.659	$\text{var}(e_{13})$	4.246	1.564	2.714
$\lambda_{T_{23}}$	8.745	0.750	11.665	$\text{var}(e_{23})$	9.706	2.037	4.764
$\lambda_{T_{33}}$	6.753	0.657	10.281	$\text{var}(e_{33})$	12.474	1.882	6.627
$\lambda_{M_{11}}$	10.201	0.574	17.771	$\text{cor}(f_{T_1}, f_{T_2})$	0.915	0.015	61.848
$\lambda_{M_{21}}$	9.563	0.507	18.879	$\text{cor}(f_{T_1}, f_{T_3})$	0.864	0.022	39.085
$\lambda_{M_{31}}$	8.559	0.521	16.421	$\text{cor}(f_{T_2}, f_{T_3})$	0.858	0.023	37.862
$\lambda_{M_{12}}$	10.082	0.679	14.859				
$\lambda_{M_{22}}$	10.249	0.720	14.236				
$\lambda_{M_{32}}$	8.695	0.658	13.223				
$\lambda_{M_{13}}$	9.058	0.566	15.991				
$\lambda_{M_{23}}$	8.585	0.627	13.687				
$\lambda_{M_{33}}$	7.821	0.508	15.385				

¹以後、1 番目の方法を“自己評価”，3 番目の方法を“上司評価”，3 番目の方法を“同僚評価”とする。

²研究 II と研究 III では、いずれも同一のデータに対し CT-UM モデルを適用している。表 3.2 に示した推定値と、表 5.2 に示した推定値が異なるのは、モデルを推定する際の、識別のための制約の置き方が異なるためであり、いずれの場合でも標準化解は一致している。研究 II の表 3.2 には、モデル識別のための制約として、同じ因子からパスを受けている観測変数のなかから任意の 1 つを選び、その因子負荷量を 1 に固定する方法を用いたときの非標準化解を示した。研究 III では、MCMC 法による推定結果や、多母集団同時解析モデルの適用結果と比較検討を行うことを考慮して、各因子の分散を 1 に固定することでモデルを識別させる方法を選択した。本研究では、Mplus (ver.7.11) を用いて CT-UM モデルの推定を実行した。

表 5.2 の結果をもとに，(5.8) 式，(5.9) 式，(5.10) 式に従って各変数ごとに計算した信頼性係数と妥当性に関する係数は表 5.3 の通りであった³。

表 5.3 信頼性係数と妥当性係数

変数	Rel(x_{ij})	Con(x_{ij})	Dis(x_{ij})
x_{11}	0.815	0.256	0.559
x_{21}	0.852	0.262	0.591
x_{31}	0.751	0.256	0.495
x_{12}	0.936	0.420	0.516
x_{22}	0.903	0.426	0.477
x_{32}	0.893	0.459	0.435
x_{13}	0.971	0.408	0.563
x_{23}	0.939	0.478	0.461
x_{33}	0.895	0.382	0.513

“自己評価”の信頼性は低いものの，信頼性はほとんどの変数で 0.8 以上の値となっていることから，測定の安定性の高さが示された。一方で，収束的妥当性係数よりも非弁別的妥当性係数が大きい傾向にあり，生データの相関行列からも示唆された通り，妥当性が低いデータであることが確認できた。特に，“自己評価”において収束的妥当性が低いことがわかる。

続いて，本研究で提案する MCMC 法を用いた結果について以下に示す。推定には，Stan(Ver 2.4.0) および統計解析環境 R に実装されているパッケージ rstan を利用した。単一の連鎖から 5000 回のサンプリングを行い，500 回をバーンイン期間として破棄し，4500 個の MCMC 標本を推定に用いた。推定結果および，それをもとに計算した信頼性と妥当性に関する係数を表 5.4 に示した。

Gelman & Rubin(1992) で提案された収束判定指標 \hat{R} を参照することで，収束判定を行った。 \hat{R} の値は 1 に近いほど収束状況が良好であることを意味し，1.1 以下であれば収束していると判断してよいとされている。今回の推定結果では，すべての母数に関して \hat{R} は最大でも 1.002 以下であったため，マルコフ連鎖の

³研究 II と研究 III では，モデル識別のための制約の置き方が異なるものの，同一のデータに対し CT-UM モデルを適用している。そのため，推定値を利用して計算される信頼性係数および収束的妥当性係数，非弁別的妥当性係数の値はいずれの場合も同じになる。したがって，これらの係数は，研究 II の結果として表 3.2 に示した値の再掲である。

表 5.4 MCMC 法による推定結果と係数

変数	λ_T	λ_M	$\text{var}(e_{ij})$	$\text{Rel}(x_{ij})$	$\text{Con}(x_{ij})$	$\text{Dis}(x_{ij})$
x_{11}	6.870	10.291	34.632	0.816	0.251	0.564
x_{21}	6.204	9.731	23.318	0.851	0.246	0.605
x_{31}	6.168	8.604	37.246	0.751	0.255	0.496
x_{12}	8.645	10.433	13.110	0.933	0.380	0.553
x_{22}	9.128	10.654	22.445	0.898	0.380	0.518
x_{32}	8.462	9.011	20.835	0.880	0.412	0.468
x_{13}	7.882	8.976	9.570	0.937	0.408	0.529
x_{23}	9.117	8.426	21.534	0.877	0.473	0.404
x_{33}	7.074	7.710	29.111	0.790	0.361	0.429

収束が示唆された。また、誤差分散に関する MCMC 標本の時系列プロットを図 5.1 に示した。図 5.1 から、マルコフ連鎖の収束が良好であることが確認できる。

多母集団同時解析モデルによる推定結果（非標準化解）および、それをもとに計算した信頼性係数と妥当性に関する係数を表 5.5 に示した。なお、多母集団同時解析モデルにおいては、群の構成員が 1 人の場合には推定ができないため、6 人からの“同僚評価”が得られていた 1 つのオブザベーションは除いた上でデータを 5 つの群に分け、分析を実行した（表 5.1 参照）。適合度指標の値は、CFI=0.976, TLI=0.979, RMSEA=0.074 であった⁴。

表 5.4 に示した MCMC 法を用いた推定結果では、表 5.3 に示した従来の方法による結果と比較して、“同僚評価”の誤差分散と、“自己評価”・“上司評価”の誤差分散との差が小さくなっている。表 5.5 の多母集団同時解析による結果でもまた、従来の方法よりも“同僚評価”の誤差分散が大きく推定され、他の変数の誤差分散と値が近くなっている。

続いて、MCMC 法によるアプローチを適用することで、信頼性係数および妥当性に関する係数にどのような変化が見られるか、表 5.3 と表 5.4 の比較を行う。信頼性係数の値は、“自己評価”および“上司評価”ではほとんど変化がなかったが、MCMC 法を用いることで“同僚評価”の信頼性係数の値が低く算出された。

⁴分析には、Mplus (ver.7.11) を用いた。

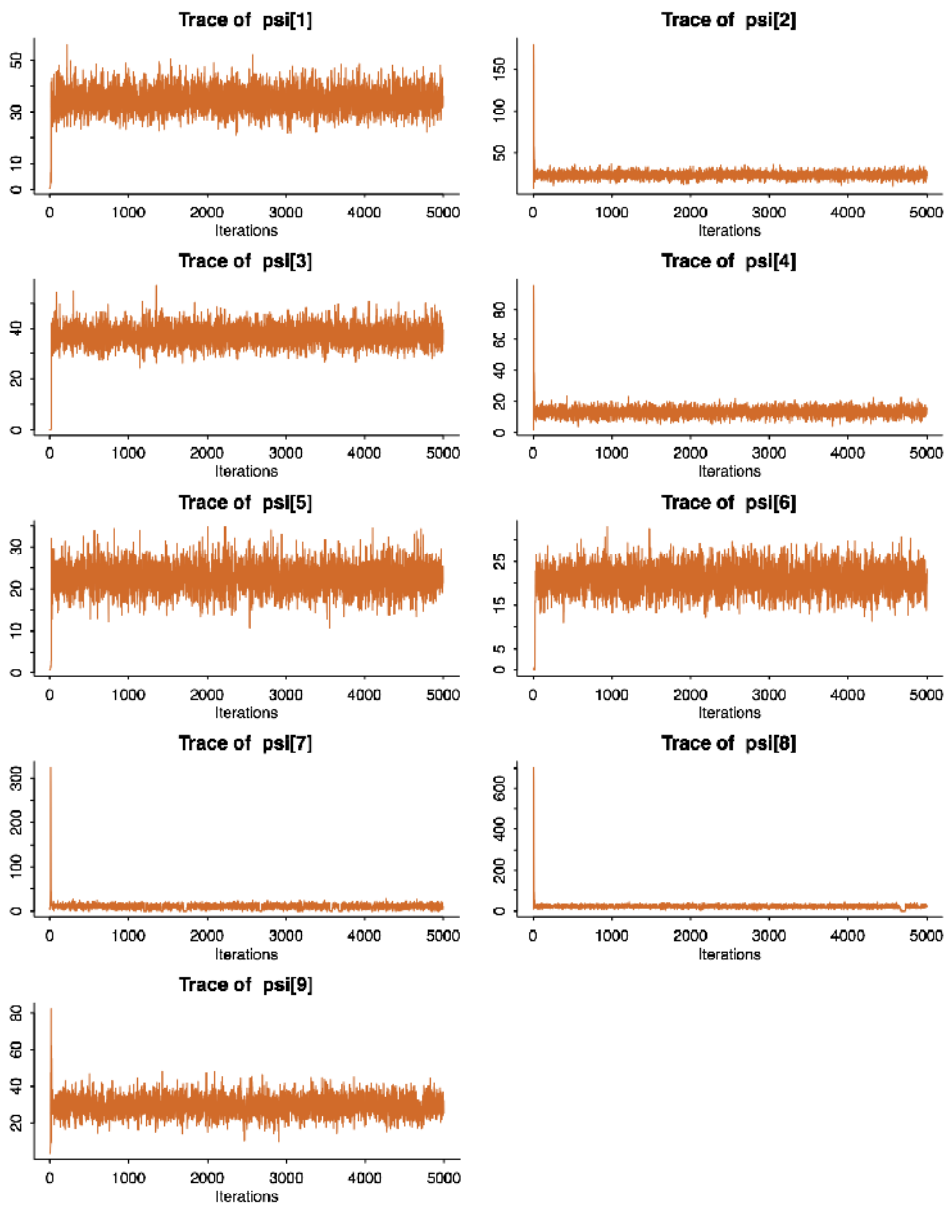


図 5.1 誤差分散の時系列プロット

表 5.5 多母集団同時解析モデルの推定結果と係数

変数	λ_T	λ_M	$\text{var}(e_{ij})$	$\text{Rel}(x_{ij})$	$\text{Con}(x_{ij})$	$\text{Dis}(x_{ij})$
x_{11}	6.393	10.474	31.930	0.825	0.224	0.601
x_{21}	6.355	9.306	25.279	0.834	0.265	0.569
x_{31}	5.942	8.634	34.879	0.759	0.244	0.515
x_{12}	8.816	10.114	12.426	0.935	0.404	0.531
x_{22}	9.537	10.041	21.887	0.898	0.425	0.472
x_{32}	8.334	8.909	20.672	0.878	0.410	0.468
x_{13}	8.006	8.715	12.571	0.918	0.420	0.498
x_{23}	8.823	8.698	22.990	0.970	0.441	0.429
x_{33}	7.184	7.545	26.453	0.804	0.382	0.422

特に、 x_{33} については、約 0.1 の低下が見られ、信頼性係数の値は 0.7 台にまで下がった。これは、因子負荷量の値が従来の方法と比較してもほとんど変わらなかったのに対して、係数を計算する際の分母を構成する誤差分散の値が大きく推定されたためである。収束的妥当性係数の値には、従来の方法と提案手法との間で大きな変化はなかった。一方で、“同僚評価”の非弁別的妥当性係数は、MCMC 法によるアプローチの結果において従来の方法より小さな値となった。信頼性係数と妥当性係数に関するこれらの傾向は、従来の方法と多母集団同時解析モデルの適用結果（表 5.5）との比較においても同様に観察される。

5.4 研究 III 総合考察

本研究では、同一立場内に複数の評価者がいる場合の MTMR データについて、その平均値を用いた確認的因子分析モデルにおける誤差分散を数学的に導出した。それを踏まえ、確認的因子分析の枠組みから評価者の人数を適切に評価し、測定の信頼性と妥当性を検討することを目的として、MCMC 法によるアプローチの提案を行い、適用例では実データを用いて提案手法の有効性を確認した。

適用例では、同一立場の評価者から複数の評定結果が得られた場合に、平均値を利用したデータに従来のモデルをそのまま適用すると、誤差分散が過小評

価されることが明らかとなった。この過小評価は、提案モデルの適用により改善される。さらに、同一立場内に複数の評価者が存在する MTMR データに対して、MCMC 法を用いて測定誤差を正しく推定することで、その結果をもとに求められる信頼性係数と妥当性係数の値もまた変化することが示された。したがって、誤差分散の値を過小評価してしまう従来の方法によって、測定の信頼性と妥当性を検討する場合には、誤った解釈につながる可能性があることが示唆された。企業における 360 度フィードバックのように、測定の信頼性と妥当性がそのアセスメントの有効性に直接関わるような場面において、信頼性と妥当性を正確に評価することの重要性は大きいだろう。MCMC 法による推定を適用することで、被評価者ごとに異なる他者評価の人数を統計的に適切に処理した上で、正確な信頼性と妥当性を求められることが示された。

適用例では、MCMC 法によるアプローチに加え、多母集団同時解析モデルによるアプローチの結果も示した。これらの結果から、多母集団同時解析モデルでもまた、誤差分散の過小評価という問題を改善できる可能性が示唆された。しかしながら、今回のデータのように、評価者の数によって群分けを行った際に構成員が 1 人しかいない群については、推定ができなくなってしまうため、その群を構成するオブザベーションはデータから除外するという対処をとらざるを得ない。一方で、MCMC 法によるアプローチでは、事前に群分けしたデータを作成する必要もなく、特定の数人の評価者を有するオブザベーションが 1 人しかいない場合でも、そのデータを含めたすべてのデータを用いて推定を行うことができるというメリットがある。

また、今回の適用例で用いたデータでは、同僚評価のみ複数の評価者からの評価が得られている状況であったが、これに加えて、例えば部下評価についても複数の評価者からの評価が得られるという状況も考えられる。このように、複数の評価者からの評価が得られている評価者の立場が 2 つ以上となった場合に、多母集団同時解析によるアプローチを適用しようとする、データの群分けが煩雑になるという問題が生じる。MCMC 法によるアプローチではデータを群分けすることなく、従来の方法とまったく同じデータセットを利用して分析を実行することが可能であるため、この問題も回避することができる。

第6章 総合考察

本論文では、確認的因子分析モデルの枠組みで MTMM データを分析し、測定信頼性および収束的妥当性、弁別的妥当性を定量的に検討するための方法について論じてきた。まず、研究Iでは、汎用的かつより実用的な MTMM データのための確認的因子分析モデルの開発を目的として、方法因子の因子得点の和を 0 に制約することで、信頼性および妥当性の解釈が 1 つのデータに対して必ず一意に定まるモデルを提案した。先行研究によって提案されてきたこれまでのモデルは、識別不定や不適解に陥りやすいといった分析面での問題や、1 つのデータに対して解が一意に定まらないという解釈面での問題から、実用性に欠けると考えられる。これに対して、提案モデルによる分析結果は必ず一意に定まり、また、シミュレーション研究を通して、提案モデルは識別される可能性が非常に高いことも確認できた。

ここで改めて、方法因子の因子得点の和を 0 とする制約は、決して便宜的なものではないという点を主張しておきたい。MTMM データの心理学的分析において、分析者の主たる興味の対象は、決して方法ではなく、特性の方である。なぜなら、測定したい構成概念は特性そのものであり、その測定方法は、収束的妥当性と弁別的妥当性の検討するための補助的道具に過ぎない。このとき、方法因子の因子得点の和を 0 に制約することは、測定を意図している特性に対して方法の影響は誤差の一部と見なすという分析者の仮定を表現するものであり、理に適っていると言える。したがって、提案モデルは十分に実用に供するモデルであると結論付けてよいであろう。

一方で、提案モデルの識別に関しては、シミュレーション研究による確認にとどまっており、必ず識別されることが証明されたわけではない。今後、提案モデルの有用性の証左を示すためにも、モデルが識別可能であることを数学的に証明する必要がある。

研究Iは、いかなる MTMM データに対しても汎用的なモデルを提案することを目的とした研究であった。研究Iにおいて様々な MTMM 行列の分析を行った結果から、“方法”として何を用いるかによって、測定の信頼性と妥当性に与える影響に系統的な違いがあることが示唆された。特に、評価者の違いを“方法”とした MTMR データでは、安定した測定を実現することは難しく、信頼性、妥当性ともに低くなる傾向が確認された。そこで、研究IIおよび研究IIIでは、MTMR データの分析に焦点を絞り、議論を進めた。MTMM データは、収束的妥当性および弁別的妥当性の検討のために有用であるものの、学術目的の調査等で、実際に、同一の特性を複数の測定方法によって測定するという機会はほとんどない。そのような状況において、人事アセスメントの現場では、MTMR データの収集が比較的盛んに行われている。すなわち、我が国においても大企業を中心に実施されている 360 度フィードバックである。360 度フィードバックを中心とした MTMR データは、現実場面でもっとも扱う機会が多い MTMM 形式のデータと言えよう。360 度フィードバックの結果として得られる MTMR データを分析対象とした信頼性と妥当性の検討は、人事アセスメントの質を吟味するという意味で非常に重要であり、現場からの需要も高いと考えられる。

研究IIでは、立場の異なる複数の評価者からの評定結果を特性ごとに合計し、各被評価者の得点(特性値)を計算するという場面に着目した。この合計得点の信頼性と妥当性という観点から、評価者ごとに適切な項目配分を決定する方法について提案した。研究Iでは、Eid (2000) に基づき、信頼性と妥当性について定量的に表現するための一貫性係数、方法特異性係数を定義した。しかし、これらの係数の意味づけは複雑で、観測変数の一貫性係数と真値の一貫性係数など、区別が難しい概念が併存しており、解釈が容易ではなかった。そこで、研究II以降では、収束的妥当性と弁別的妥当性に関する係数を、より簡潔かつ明快な解釈ができるよう再定義した。

合計得点の信頼性と妥当性という観点から得られる項目配分に関する示唆は、これまで蓄積されてきた人事アセスメントの知見に照らしても納得できる結果となった。しかしながら、実際に項目を増やしたり減らしたりする際に、どの項目を除くのか、あるいはどのような項目を加えるべきかという判断までを、本研究の範囲で行うことはできない。研究IIの方法によって得られた示唆に基づ

き、実際に既存のアセスメントツールから項目数を変化させるためには、内的整合性の観点からの項目分析や、項目プールの用意など、付加的な検討事項を要する。

最後に研究 III では、同一立場内に複数の評価者が存在する MTMR データの分析において、評価者の人数を適切に考慮し、信頼性と妥当性を検討するための方法を提案した。推定に HMC 法を導入した MCMC 法を用いることで、評価者の人数に応じて被評価者ごとに異なる測定誤差の推定を可能にした。HMC 法による MCMC 法を用いた推定は、統計解析環境 R のパッケージ `rstan` および `Stan` を利用して容易に実行することができる。標本サイズが小さくても安定的な推定が可能であり、複雑なモデルも柔軟に表現することができる MCMC 法は、今後広く利用されることが期待される。また、本研究により、誤差分散が正しく推定されることで、分散成分の分解によって定義される信頼性係数および、収束的妥当性と弁別的妥当性に関する係数もまた、評価者の人数の影響を反映したより正確な値となるはずである。したがって、これらの係数を利用することで、測定の信頼性と妥当性に関して正しく解釈することが可能となる。

Campbell & Fiske (1959) を嚆矢として、心理学的測定においても弁別的妥当性と収束的妥当性を検討するために MTMM 行列が利用されるようになり、現在、その分析には確認的因子分析モデルが広く用いられている。しかしながらこれまで、モデルの適用から、信頼性と妥当性の解釈までを手続き化して示した先行研究はほとんどなく、確認的因子分析モデルの推定結果を用いて定量的に信頼性と妥当性に関する係数を定義する試みは Eid (2000) が初めてであろう。本論文では、3つの研究を通して、モデルの適用から、信頼性と妥当性を解釈するための係数の算出、および、それらの解釈までを一貫して行ってきた。MTMM データを用いて測定の信頼性と収束的妥当性、弁別的妥当性を検討するための手続きを体系化して示したことで、本研究の成果が様々な研究の現場で実際に活用される日が来ることを願う。

また、MTMM データはその有用性にもかかわらず、データの性質上、個人の研究者による調査の実施は難しく、企業における 360 度フィードバック以外ではほとんど収集されていないのが現実である。MTMR データに限定すれば、例えば、児童の人格特性や行動傾向に関する調査で、本人に加えて親や学校の先

生に回答を求める場合にも用いることができるだろう。今後、人事アセスメントの現場ばかりではなく、心理学の研究においても MTMM データを用いた妥当性研究の有用性の認知が広まり、MTMM データの適用場面が増えることを期待する。

Coenders & Saris (2000)

1.000
0.464 1.000
0.340 0.223 1.000
0.574 0.241 0.330 1.000
0.349 0.762 0.195 0.300 1.000
0.175 0.019 0.646 0.390 0.119 1.000
0.639 0.346 0.309 0.630 0.326 0.281 1.000
0.399 0.788 0.166 0.251 0.791 0.052 0.420 1.000
0.276 0.121 0.657 0.282 0.143 0.692 0.445 0.237 1.000

Kelly & Fiske (1951)

1.00
0.37 1.00
-0.24 -0.14 1.00
0.25 0.46 0.08 1.00
0.35 0.19 0.09 0.31 1.00
0.71 0.35 -0.18 0.26 0.41 1.00
0.39 0.53 -0.15 0.38 0.29 0.37 1.00
-0.27 -0.31 0.43 -0.06 0.03 -0.15 -0.19 1.00
0.03 -0.05 0.03 0.20 0.07 0.11 0.23 0.19 1.00
0.19 0.05 0.04 0.29 0.47 0.33 0.22 0.19 0.29 1.00
0.48 0.31 -0.22 0.19 0.12 0.46 0.36 -0.15 0.12 0.23 1.00
0.17 0.42 -0.10 0.10 -0.03 0.09 0.24 -0.25 -0.11 -0.03 0.23 1.00
-0.04 -0.13 0.22 -0.13 -0.05 -0.04 -0.11 0.31 0.06 0.06 -0.06 -0.12 1.00
0.13 0.27 -0.03 0.22 -0.04 0.10 0.15 0.00 0.14 -0.03 0.16 0.26 0.11 1.00
0.37 0.15 -0.22 0.09 0.26 0.27 0.12 -0.07 0.05 0.35 0.21 0.15 0.17 0.31 1.00

Kothandapani (1971)

1.00
0.26 1.00
0.19 0.43 1.00
0.58 0.34 0.05 1.00
0.20 0.50 0.22 0.66 1.00
-0.03 0.08 0.49 0.31 0.65 1.00
0.49 0.06 -0.06 0.60 0.20 0.04 1.00
0.20 0.28 0.09 0.41 0.44 0.24 0.48 1.00
0.18 -0.05 0.42 0.05 0.07 0.47 0.28 0.32 1.00
0.45 0.20 0.12 0.59 0.41 0.31 0.50 0.30 0.13 1.00
0.25 0.39 0.32 0.31 0.51 0.39 0.13 0.35 0.13 0.65 1.00
0.17 0.12 0.49 0.15 0.34 0.60 0.14 0.24 0.50 0.55 0.61 1.00

Levin (1973)

1.00											
0.50	1.00										
0.41	0.48	1.00									
0.67	0.59	0.40	1.00								
0.45	0.33	0.26	0.55	1.00							
0.36	0.32	0.31	0.43	0.72	1.00						
0.25	0.21	0.25	0.30	0.59	0.72	1.00					
0.46	0.36	0.28	0.51	0.85	0.80	0.69	1.00				
0.53	0.41	0.34	0.56	0.71	0.58	0.43	0.72	1.00			
0.50	0.45	0.29	0.52	0.59	0.55	0.42	0.63	0.84	1.00		
0.36	0.30	0.28	0.37	0.53	0.51	0.43	0.57	0.62	0.57	1.00	
0.52	0.43	0.31	0.59	0.68	0.60	0.46	0.73	0.92	0.89	0.63	1.00

Maas, Lensvelt-Mulders & Hox (2009)

1.00											
0.359	1.00										
0.240	0.299	1.00									
0.837	0.284	0.179	1.00								
0.254	0.862	0.249	0.298	1.00							
0.196	0.208	0.897	0.223	0.252	1.00						
0.832	0.330	0.190	0.851	0.293	0.197	1.00					
0.247	0.833	0.250	0.272	0.864	0.225	0.361	1.00				
0.169	0.204	0.848	0.173	0.218	0.864	0.208	0.254	1.00			

Manson, Levine & Brannick (2000)

1.00											
0.90	1.00										
0.54	0.43	1.00									
0.18	0.18	0.13	1.00								
0.81	0.77	0.57	0.29	1.00							
0.74	0.75	0.27	0.08	0.75	1.00						
0.57	0.47	0.85	0.11	0.68	0.44	1.00					
0.17	0.17	0.06	0.66	0.36	0.36	0.16	1.00				
0.84	0.83	0.48	0.12	0.85	0.71	0.54	0.16	1.00			
0.86	0.84	0.48	0.16	0.75	0.73	0.50	0.20	0.88	1.00		
0.55	0.44	0.90	0.01	0.56	0.33	0.85	0.05	0.54	0.52	1.00	
0.12	0.15	0.19	0.78	0.32	0.08	0.16	0.60	0.24	0.17	0.11	1.00

Muis, Winne & Jamieson-Noel (2007)

1.00
0.27 1.00
0.04 0.35 1.00
0.23 0.44 0.43 1.00
0.32 0.32 0.59 0.46 1.00
0.40 0.59 0.31 0.39 0.47 1.00
0.41 0.49 0.36 0.64 0.49 0.48 1.00
0.22 0.65 0.15 0.39 0.20 0.52 0.42 1.00
0.29 0.38 0.26 0.52 0.37 0.38 0.66 0.34 1.00
0.39 0.60 0.47 0.58 0.50 0.54 0.66 0.42 0.67 1.00
0.29 0.40 0.27 0.56 0.28 0.34 0.55 0.31 0.51 0.70 1.00
0.19 0.48 0.30 0.50 0.37 0.42 0.57 0.41 0.69 0.66 0.54 1.00

Ostrom (1969)

1.00
0.57 1.00
0.63 0.62 1.00
0.71 0.59 0.68 1.00
0.67 0.67 0.71 0.79 1.00
0.69 0.62 0.72 0.79 0.81 1.00
0.54 0.39 0.49 0.58 0.51 0.56 1.00
0.59 0.61 0.58 0.60 0.69 0.60 0.43 1.00
0.63 0.50 0.63 0.69 0.67 0.71 0.49 0.56 1.00
0.65 0.63 0.69 0.72 0.73 0.72 0.54 0.66 0.65 1.00
0.61 0.51 0.56 0.62 0.68 0.64 0.53 0.52 0.56 0.76 1.00
0.56 0.48 0.60 0.63 0.66 0.66 0.46 0.48 0.62 0.77 0.68 1.00

Roshal, Frieze & Wood (1971)

1.00
0.68 1.00
0.11 0.26 1.00
0.30 0.27 0.01 1.00
0.18 0.23 0.12 0.64 1.00
0.00 0.07 0.20 0.34 0.56 1.00
0.28 0.26 0.11 0.42 0.46 0.27 1.00
0.19 0.21 0.23 0.36 0.48 0.37 0.70 1.00
0.00 0.03 0.20 0.04 0.10 0.29 0.09 0.30 1.00

研究Iで使用したEQSのコード

例として、Coenders & Saris (2000) の分析に使用したEQSのコードを以下に記載する。

提案モデル

```
/TITLE
Coenders&Saris(2000):提案モデル
/SPECIFICATIONS
VAR=9; CAS=406; MAT=COR; ANA=COR;
/EQUATIONS
V1 = F1 + F4 + E1;
V2 = F2 + *F4 + E2;
V3 = F3 + *F4 + E3;
V4 = *F1 + F5 + E4;
V5 = *F2 + *F5 + E5;
V6 = *F3 + *F5 + E6;
V7 = *F1 + F6 + E7;
V8 = *F2 + *F6 + E8;
V9 = *F3 + *F6 + E9;
/VARIANCES
F1 TO F6 = *;
/COVARIANCES
F1,F2=*; F1,F3=*; F2,F3=*;
F4,F5=*; F4,F6=*; F5,F6=*;
/CONSTRAINTS
(F4, F4) + 2(F4,F5) + (F5, F5) - (F6, F6) = 0;
(F4,F6) + (F4,F4) + (F4,F5) = 0;
(F5,F6) + (F5,F5) + (F4,F5) = 0;

/MATRIX
1.000
0.464 1.000
0.340 0.223 1.000
0.574 0.241 0.330 1.000
0.349 0.762 0.195 0.300 1.000
0.175 0.019 0.646 0.390 0.119 1.000
0.639 0.346 0.309 0.630 0.326 0.281 1.000
0.399 0.788 0.166 0.251 0.791 0.052 0.420 1.000
0.276 0.121 0.657 0.282 0.143 0.692 0.445 0.237 1.000
/TECHNICAL
ITR=50;

/END
```

CT-CMモデル

```
/TITLE
Coenders&Saris(2000):CT-CM
/SPECIFICATIONS
VAR=9; CAS=406; MAT=COR; ANA=COR;
/EQUATIONS
V1 = F1 + F4 + E1;
V2 = F2 + *F4 + E2;
V3 = F3 + *F4 + E3;
V4 = *F1 + F5 + E4;
V5 = *F2 + *F5 + E5;
V6 = *F3 + *F5 + E6;
V7 = *F1 + F6 + E7;
V8 = *F2 + *F6 + E8;
V9 = *F3 + *F6 + E9;
/VARIANCES
F1 TO F6 = *;
/COVARIANCES
F1,F2=*; F1,F3=*; F2,F3=*;
F4,F5=*; F4,F6=*; F5,F6=*;
---相関行列の入力は提案モデルと同じなので以下省略---
/END
```

CT-C(M-1)モデル (M1を基準とした場合)

```
/TITLE
Coenders&Saris(2000):CT-C(M-1)
/SPECIFICATIONS
VAR=9; CAS=406; MAT=COR; ANA=COR;
/EQUATIONS
V1 = F1 + E1;
V2 = F2 + E2;
V3 = F3 + E3;
V4 = *F1 + F5 + E4;
V5 = *F2 + *F5 + E5;
V6 = *F3 + *F5 + E6;
V7 = *F1 + F6 + E7;
V8 = *F2 + *F6 + E8;
V9 = *F3 + *F6 + E9;
/VARIANCES
F1 TO F6 = *;
E1 TO E9 = *;
/COVARIANCES
F1,F2=*; F1,F3=*; F2,F3=*;
F5,F6=*;
---相関行列の入力は提案モデルと同じなので以下省略---
/END
```

研究Iのシミュレーションに使用したRコード

```
set.seed(135)
library(MASS)
par.t <- seq(0.3, 0.7, 0.2)
par.m <- seq(0.2, 0.4, 0.1)
par.rt1 <- seq(0.3, 0.7, 0.2)
par.rt2 <- seq(0.3, 0.7, 0.2)
par.rt3 <- seq(0.3, 0.7, 0.2)
for(i1 in 1:3){
  for(i2 in 1:3){
    for(i3 in 1:3){
      for(i4 in 1:3){
        for(i5 in 1:3){
          for(i6 in 1:3){
            for(i7 in 1:3){
              for(i8 in 1:3){
                for(i9 in 1:3){
                  i1 <- i2 <- i3 <- i4 <- i5 <- i6 <- i7 <- i8 <- i9 <- 2
                  Lt <- cbind(rep(c(par.t[i1], 0, 0), 3),
                                rep(c(0, par.t[i2], 0), 3),
                                rep(c(0, 0, par.t[i3]), 3))
                  Lm <- cbind(c(rep(par.m[i4], 3), 0, 0, 0, 0, 0, 0),
                                c(0, 0, 0, rep(par.m[i5], 3), 0, 0, 0),
                                c(0, 0, 0, 0, 0, 0, rep(par.m[i6], 3)))
                  Pt1 <- rnorm(1, mean = par.rt1[i7], sd = 0.02)
                  Pt2 <- rnorm(1, mean = par.rt2[i8], sd = 0.02)
                  Pt3 <- rnorm(1, mean = par.rt3[i9], sd = 0.02)
                  #特性因子間相関
                  St <- matrix(c(
                    1, Pt1, Pt2,
                    Pt1, 1, Pt3,
                    Pt2, Pt3, 1), 3, 3)
                  #方法因子間相関
                  k <- matrix(c(1, 0, -1, 0, 1, -1), 3, 2)
                  Sm <- k %*% matrix(c( 1, -0.5, -0.5, 1), 2, 2) %*% t(k)
                  P <- Lt %*% St %*% t(Lt) + Lm %*% Sm %*% t(Lm)
                  E <- diag(diag(diag(9) - P), 9, 9)
                  R <- P + E

                  name.cond <- paste(
                    "Lt1=", par.t[i1], ",", "Lt2=", par.t[i2], ",",
                    "Lt3=", par.t[i3], ",",
                    "Lm1=", par.m[i4], ",", "Lm2=", par.m[i5], ",",
                    "Lm3=", par.m[i6], ",",
                    "Pt1=", par.rt1[i7], ",", "Pt2=", par.rt1[i8], ",",
                    "Pt3=", par.rt1[i9],
```

```

".csv", sep="")

write.csv(R,
paste("C:/Users/tyosem/Desktop/cormat1/", name.cond, sep=""),
row.names=F, fileEncoding="SJIS")
}
}
}
}
}
}
}
}
}

list.cmat <- list.files("/Volumes/Orange/cormat/")
names.cmat <- sub(".csv", "", list.cmat)
cmat <- lapply(paste("/cormat/", list.cmat, sep=""), read.csv)

names(cmat) <- names.cmat
for(i in 1:length(cmat)){
cor <- cmat[[i]]
}

library(lavaan)
simu <- function(N, l_t, l_m, corT, corM, dat=FALSE){
  Lt <- rbind(diag(l_t,3,3), diag(l_t,3,3), diag(l_t,3,3))
  Lm <- cbind(c(l_m,l_m,l_m,0,0,0,0,0,0),c(0,0,0,l_m,l_m,l_m,0,0,0),
             c(0,0,0,0,0,0,l_m,l_m,l_m))
  L <- cbind(Lt, Lm)
  m <- numeric(6)
  #因子間相關
  S <- matrix(c(1, corT, corT, 0, 0, 0,
               corT, 1, corT, 0, 0, 0,
               corT, corT, 1, 0, 0, 0,
               0, 0, 0, 1, corM, corM,
               0, 0, 0, corM, 1, corM,
               0, 0, 0, corM, corM, 1), 6, 6)

  #set.seed(s)
  fac <- t(mvrnorm(N, mu=m, Sigma=S))
  #set.seed(s)
  e <- matrix(rnorm(9*N, sd=1), 9, N)
  x <- t(L %*% fac + e)
  X <- cor(x)
  colnames(X) <- rownames(X) <- paste(c("x14", "x24", "x34",
                                         "x15", "x25", "x35", "x16", "x26", "x36"))
}

```

```

MTMM_0 <- ' f1 =~ x14 + x15 + x16
           f2 =~ x24 + x25 + x26
           f3 =~ x34 + x35 + x36
           f4 =~ 1*x14 + x24 + x34 + (-1)*x16 +
                 start(-1.0)*x26 + start(-1.0)*x36
           f5 =~ 1*x15 + x25 + x35 + (-1)*x16 +
                 start(-1.0)*x26 + start(-1.0)*x36
           f1 ~~ 0*f4 + 0*f5
           f2 ~~ 0*f4 + 0*f5
           f3 ~~ 0*f4 + 0*f5 '
MTMM_0.fit <- cfa(MTMM_0, sample.cov=X, sample.nobs=N)
if(dat){
  summary(MTMM_0.fit)
return(X)

}else
return(MTMM_0.fit@Fit@converged)
}

kakka01 <- function(ite){
  rslt <- numeric(ite)
  for(i in 1:ite){
    rslt[i] <- simu(N=400, l_t=0.9, l_m=0.5, corT=0.5, corM=0.1)
  }
  return(list(rslt, sum(rslt)))
}

kakka02 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, l_t=0.9, l_m=0.5, corT=0.5, corM=0.2)
  }
  return(list(rslt, sum(rslt)))
}

kakka03 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.5, corT=0.3, corM=0.1)
  }
  return(list(rslt, sum(rslt)))
}

kakka04 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){

```



```

rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.5, corT=0.3, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

kakka05 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.5, corT=0.1, corM=0.1)
}
return(list(rslt, sum(rslt)))
}

kakka06 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.5, corT=0.1, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

kakka07 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.5, corT=0.5, corM=0.1)
}
return(list(rslt, sum(rslt)))
}

kakka08 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.5, corT=0.5, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

kakka09 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.5, corT=0.3, corM=0.1)
}
return(list(rslt, sum(rslt)))
}

kakka10 <- function(ite){
rslt <- numeric(ite)

```

```

for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.5, corT=0.3, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

kakka11 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.5, corT=0.1, corM=0.1)
}
return(list(rslt, sum(rslt)))
}

kakka12 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.5, corT=0.1, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

kakka13 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.5, corT=0.5, corM=0.1)
}
return(list(rslt, sum(rslt)))
}

kakka14 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.5, corT=0.5, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

kakka15 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.5, corT=0.3, corM=0.1)
}
return(list(rslt, sum(rslt)))
}

kakka16 <- function(ite){

```

```

rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.5, corT=0.3, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

kakka17 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.5, corT=0.1, corM=0.1)
}
return(list(rslt, sum(rslt)))
}

kakka18 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.5, corT=0.1, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

kakka19 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.3, corT=0.5, corM=0.1)
}
return(list(rslt, sum(rslt)))
}

kakka20 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.3, corT=0.5, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

kakka21 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.3, corT=0.3, corM=0.1)
}
return(list(rslt, sum(rslt)))
}

```

```

kakka22 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.3, corT=0.3, corM=0.2)
  }
  return(list(rslt, sum(rslt)))
}

kakka23 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.3, corT=0.1, corM=0.1)
  }
  return(list(rslt, sum(rslt)))
}

kakka24 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.3, corT=0.1, corM=0.2)
  }
  return(list(rslt, sum(rslt)))
}

kakka25 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.3, corT=0.5, corM=0.1)
  }
  return(list(rslt, sum(rslt)))
}

kakka26 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.3, corT=0.5, corM=0.2)
  }
  return(list(rslt, sum(rslt)))
}

kakka27 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.3, corT=0.3, corM=0.1)
  }
  return(list(rslt, sum(rslt)))
}

```

```

kakka28 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.3, corT=0.3, corM=0.2)
  }
  return(list(rslt, sum(rslt)))
}

kakka29 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.3, corT=0.1, corM=0.1)
  }
  return(list(rslt, sum(rslt)))
}

kakka30 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.3, corT=0.1, corM=0.2)
  }
  return(list(rslt, sum(rslt)))
}

kakka31 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.3, corT=0.5, corM=0.1)
  }
  return(list(rslt, sum(rslt)))
}

kakka32 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.3, corT=0.5, corM=0.2)
  }
  return(list(rslt, sum(rslt)))
}

kakka33 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.3, corT=0.3, corM=0.1)
  }
  return(list(rslt, sum(rslt)))
}

```

```

}

kakka34 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.3, corT=0.3, corM=0.2)
  }
  return(list(rslt, sum(rslt)))
}

kakka35 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.3, corT=0.1, corM=0.1)
  }
  return(list(rslt, sum(rslt)))
}

kakka36 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.3, corT=0.1, corM=0.2)
  }
  return(list(rslt, sum(rslt)))
}

kakka37 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.1, corT=0.5, corM=0.1)
  }
  return(list(rslt, sum(rslt)))
}

kakka38 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.1, corT=0.5, corM=0.2)
  }
  return(list(rslt, sum(rslt)))
}

kakka39 <- function(ite){
  rslt <- numeric(ite)
  for(k in 1:ite){
    rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.1, corT=0.3, corM=0.1)
  }
}

```

```

return(list(rslt, sum(rslt)))
}

kakka40 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.1, corT=0.3, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

kakka41 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.1, corT=0.1, corM=0.1)
}
return(list(rslt, sum(rslt)))
}

kakka42 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.9, l_m=0.1, corT=0.1, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

kakka43 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.1, corT=0.5, corM=0.1)
}
return(list(rslt, sum(rslt)))
}

kakka44 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.1, corT=0.5, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

kakka45 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.1, corT=0.3, corM=0.1)
}

```

```

}
return(list(rslt, sum(rslt)))
}

kakka46 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.1, corT=0.3, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

kakka47 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.1, corT=0.1, corM=0.1)
}
return(list(rslt, sum(rslt)))
}

kakka48 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.6, l_m=0.1, corT=0.1, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

kakka49 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.1, corT=0.5, corM=0.1)
}
return(list(rslt, sum(rslt)))
}

kakka50 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.1, corT=0.5, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

kakka51 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){

```



```

rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.1, corT=0.3, corM=0.1)
}
return(list(rslt, sum(rslt)))
}

```

```

kakka52 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.1, corT=0.3, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

```

```

kakka53 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.1, corT=0.1, corM=0.1)
}
return(list(rslt, sum(rslt)))
}

```

```

kakka54 <- function(ite){
rslt <- numeric(ite)
for(k in 1:ite){
rslt[k] <- simu(N=300, s=k, l_t=0.3, l_m=0.1, corT=0.1, corM=0.2)
}
return(list(rslt, sum(rslt)))
}

```

```

kakka01(100); kakka02(100); kakka03(100); kakka04(100); kakka05(100)
kakka06(100); kakka07(100); kakka08(100); kakka09(100); kakka10(100)
kakka11(100); kakka12(100); kakka13(100); kakka14(100); kakka15(100)
kakka16(100); kakka17(100); kakka18(100); kakka19(100); kakka20(100)
kakka21(100); kakka22(100); kakka23(100); kakka24(100); kakka25(100)
kakka26(100); kakka27(100); kakka28(100); kakka29(100); kakka30(100)
kakka31(100); kakka32(100); kakka33(100); kakka34(100); kakka35(100)
kakka36(100); kakka37(100); kakka38(100); kakka39(100); kakka40(100)
kakka41(100); kakka42(100); kakka43(100); kakka44(100); kakka45(100)
kakka46(100); kakka47(100); kakka48(100); kakka49(100); kakka50(100)
kakka51(100); kakka52(100); kakka53(100); kakka54(100)

```

研究IIで使用したRコード(CT-UMモデルの推定)

```
library(lavaan)
MTMMdata <- read.csv("360MTMM.csv", header=T)
T3M3data <- MTMMdata[, -1] #相関行列を計算するため, ID の列を除く
cormat <- cor(T3M3data )
round(cormat,2)
colnames(T3M3data) <- paste(c("x11", "x21", "x31",
                              "x12", "x22", "x32", "x13", "x23", "x33"))

CT3UM3 <- ' fT1 =~ x11 + x12 + x13
            fT2 =~ x21 + x22 + x23
            fT3 =~ x31 + x32 + x33
            fM1 =~ x11 + x21 + x31
            fM2 =~ x12 + x22 + x32
            fM3 =~ x13 + x23 + x33
            fT1 ~~ 0*fM1 + 0*fM2 + 0*fM3
            fT2 ~~ 0*fM1 + 0*fM2 + 0*fM3
            fT3 ~~ 0*fM1 + 0*fM2 + 0*fM3
            fM1 ~~ 0*fM2 + 0*fM3
            fM2 ~~ 0*fM3 '
```

```
T3M3.fit <- sem(CT3UM3, data=T3M3data)
summary(T3M3.fit , fit.measures=TRUE, standardized=TRUE)
```

研究IIIで使用したStanコード

```
data{
  int<lower=0> I; #オブザベーション数
  int<lower=0> J; #観測変数の数
  int<lower=0> T; #特性因子の数
  int<lower=0> M; #方法因子の数
  vector[J] y[I]; #データ
  vector<lower=1>[I] NP; #同僚人数
  vector[J] mu;
}

parameters {
  vector[18] lam; #因子負荷量
  vector<lower=0>[J] psi; #誤差分散
  corr_matrix[T] Cor_fT; #特性因子の相関行列
}

transformed parameters {
  matrix[J,T] Lam_fT; #特性因子負荷行列
  matrix[J,M] Lam_fM; #方法因子負荷行列
  vector<lower=0>[J] ppsi; #誤差分散を再定義
  matrix[J,J] Psi[I];

  for(j in 1:J){ for(t in 1:T){ Lam_fT[j,t] <- 0.0;}}
  Lam_fT[1,1] <- lam[1];
  Lam_fT[2,2] <- lam[2];
  Lam_fT[3,3] <- lam[3];
  Lam_fT[4,1] <- lam[4];
  Lam_fT[5,2] <- lam[5];
  Lam_fT[6,3] <- lam[6];
  Lam_fT[7,1] <- lam[7];
  Lam_fT[8,2] <- lam[8];
  Lam_fT[9,3] <- lam[9];
  for(j in 1:J){ for(m in 1:M){ Lam_fM[j,m] <- 0.0;}}
  Lam_fM[1,1] <- lam[10];
  Lam_fM[2,1] <- lam[11];
  Lam_fM[3,1] <- lam[12];
  Lam_fM[4,2] <- lam[13];
  Lam_fM[5,2] <- lam[14];
  Lam_fM[6,2] <- lam[15];
  Lam_fM[7,3] <- lam[16];
  Lam_fM[8,3] <- lam[17];
  Lam_fM[9,3] <- lam[18];

  #誤差分散を再定義して、同僚評定の部分のみ同僚人数で割る
  ppsi[1] <- psi[1];
  ppsi[2] <- psi[2];
}
```

```

        ppsi[3] <- psi[3];
        ppsi[4] <- psi[4];
        ppsi[5] <- psi[5];
        ppsi[6] <- psi[6];
for(i in 1:I){ #print("loop iteration: ", ppsi);
ppsi[7] <- psi[7]/NP[i];
        ppsi[8] <- psi[8]/NP[i];
        ppsi[9] <- psi[9]/NP[i];
Psi[i] <- diag_matrix(ppsi);
}
}

model {
matrix[J,J] Sigma;
#誤差分散の事前分布
psi ~ inv_gamma(0.001,0.001);

#因子負荷量の事前分布
    for (k in 1:18){
        lam[k] ~ normal(0, 1000);
    }
#因子の相関行列の事前分布 (以下を実行しなければ一様分布)
#Cor_fT ~ lkj_corr(1);
#モデル式
Sigma <- Lam_fT*Cor_fT*Lam_fT' + Lam_fM*Lam_fM';
for(i in 1:I){
y[i] ~ multi_normal(mu, Sigma+Psi[i]);
        }
}

```

研究IIIで使用したMplusのコード（多母集団同時解析）

```
TITLE: multiple group
DATA: FILE = mean_doryo.dat;
VARIABLE: NAMES = ID x11 x21 x31 x12 x22 x32 x13 x23 x33 num;
          USEVARIABLES = x11 x21 x31 x12 x22 x32 x13 x23 x33 num;
          GROUPING = num (1=g1 2=g2 3=g3 4=g4 5=g5);
ANALYSIS: TYPE=GENERAL;
          ESTIMATOR = ML;
          MODEL IS NOMEANSTRUCTURE;
MODEL: T1 by x11* x12 x13 (1-3);
       T2 by x21* x22 x23 (4-6);
       T3 by x31* x32 x33 (7-9);
       M1 by x11* x21 x31 (10-12);
       M2 by x12* x22 x32 (13-15);
       M3 by x13* x23 x33 (16-18);
       T1 with M1@0 M2@0 M3@0;
       T2 with M1@0 M2@0 M3@0;
       T3 with M1@0 M2@0 M3@0;
       M1 with M2@0 M3@0;
       M2 with M3@0;
       T1-T3@1;
       M1-M3@1;
       T1 with T2 (c1);
       T1 with T3 (c2);
       T2 with T3 (c3);
       x11 (e1);
       x21 (e2);
       x31 (e3);
       x12 (e4);
       x22 (e5);
       x32 (e6);
       x13 (e7);
       x23 (e8);
       x33 (e9); !residual variance

MODEL g2:
       T1 by x11* x12 x13 (1-3);
       T2 by x21* x22 x23 (4-6);
       T3 by x31* x32 x33 (7-9);
       M1 by x11* x21 x31 (10-12);
       M2 by x12* x22 x32 (13-15);
       M3 by x13* x23 x33 (16-18);
       T1 with M1@0 M2@0 M3@0;
       T2 with M1@0 M2@0 M3@0;
       T3 with M1@0 M2@0 M3@0;
       M1 with M2@0 M3@0;
```

```

M2 with M3@0;
T1-T3@1;
M1-M3@1;
T1 with T2 (c1);
T1 with T3 (c2);
T2 with T3 (c3);
x11 (e1);
x21 (e2);
x31 (e3);
x12 (e4);
x22 (e5);
x32 (e6);
x13 (e10);
x23 (e11);
x33 (e12);

```

MODEL g3:

```

T1 by x11* x12 x13 (1-3);
T2 by x21* x22 x23 (4-6);
T3 by x31* x32 x33 (7-9);
M1 by x11* x21 x31 (10-12);
M2 by x12* x22 x32 (13-15);
M3 by x13* x23 x33 (16-18);
T1 with M1@0 M2@0 M3@0;
T2 with M1@0 M2@0 M3@0;
T3 with M1@0 M2@0 M3@0;
M1 with M2@0 M3@0;
M2 with M3@0;
T1-T3@1;
M1-M3@1;
T1 with T2 (c1);
T1 with T3 (c2);
T2 with T3 (c3);
x11 (e1);
x21 (e2);
x31 (e3);
x12 (e4);
x22 (e5);
x32 (e6);
x13 (e13);
x23 (e14);
x33 (e15);

```

MODEL g4:

```

T1 by x11* x12 x13 (1-3);
T2 by x21* x22 x23 (4-6);
T3 by x31* x32 x33 (7-9);

```

```

M1 by x11* x21 x31 (10-12);
M2 by x12* x22 x32 (13-15);
M3 by x13* x23 x33 (16-18);
T1 with M1@0 M2@0 M3@0;
T2 with M1@0 M2@0 M3@0;
T3 with M1@0 M2@0 M3@0;
M1 with M2@0 M3@0;
M2 with M3@0;
T1-T3@1;
M1-M3@1;
T1 with T2 (c1);
T1 with T3 (c2);
T2 with T3 (c3);
x11 (e1);
x21 (e2);
x31 (e3);
x12 (e4);
x22 (e5);
x32 (e6);
x13 (e16);
x23 (e17);
x33 (e18);

```

MODEL g5:

```

T1 by x11* x12 x13 (1-3);
T2 by x21* x22 x23 (4-6);
T3 by x31* x32 x33 (7-9);
M1 by x11* x21 x31 (10-12);
M2 by x12* x22 x32 (13-15);
M3 by x13* x23 x33 (16-18);
T1 with M1@0 M2@0 M3@0;
T2 with M1@0 M2@0 M3@0;
T3 with M1@0 M2@0 M3@0;
M1 with M2@0 M3@0;
M2 with M3@0;
T1-T3@1;
M1-M3@1;
T1 with T2 (c1);
T1 with T3 (c2);
T2 with T3 (c3);
x11 (e1);
x21 (e2);
x31 (e3);
x12 (e4);
x22 (e5);
x32 (e6);
x13 (e19);

```

```
x23 (e20);  
x33 (e21);
```

```
MODEL CONSTRAINT:
```

```
e10 = (e7)/2;  
e11 = (e8)/2;  
e12 = (e9)/2;  
e13 = (e7)/3;  
e14 = (e8)/3;  
e15 = (e9)/3;  
e16 = (e7)/4;  
e17 = (e8)/4;  
e18 = (e9)/4;  
e19 = (e7)/5;  
e20 = (e8)/5;  
e21 = (e9)/5;
```

```
OUTPUT:  SAMPSTAT STANDARDIZED;  
         RESIDUAL CINTERVAL;  
         TECH4
```


引用文献

相原孝夫・南雲道朋 (2009). チームを活性化し人材を育てる 360 度フィードバック 日本経済新聞出版社.

American Psychological Association. (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, **51**(2).

American Psychological Association. (1966). *Standards for educational and psychological tests and manuals*. Washington, DC: Author.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1974). *Standards for educational and psychological tests*. Washington, DC: American Psychological Association.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Angoff, W. H. (1988). Validity: An evolving concept. In H. Wainer & H. Braun (Eds.), *Test Validity* pp. 19-32. Hillsdale, NJ: Lawrence Erlbaum.

Ashford, S. J. (1989). Self-assessment in organizations: A literature review and integrative model. *Research in Organizational Behavior*, **11**, 133-174.

馬場昌雄・馬場房子(監修), 岡村一成・小野公一(編) (2005). 産業・組織心理学 白桃書房.

Bagozzi, R. P., & Yi, Y. (1990). Assessing method variance in multitrait-multimethod matrices: The case of self-reported affect and perceptions at work. *Journal of Applied Psychology*, **75**(5), 547-560.

Bagozzi, R. P., & Yi, Y. (1992). Testing hypothesis about method, traits, and communalities in the direct product model. *Applied Psychological Measurement*, **16**, 373-380.

- Bass, B. M. (1990). *Bass & Stogdill's handbook of leadership: Theory, Research, & Managerial Applications*. New York, NY: Free Press London.
- Biesanz, J. C., & West, S. G. (2004). Towards understanding assessments of the Big Five: Multitrait-multimethod analyses of convergent and discriminant validity across measurement occasion and type of observer. *Journal of Personality*, **72**, 845-876.
- Boruch, R. F., Larkin, J. D., Wolins, L., & MacKinney, A. C. (1970). Alternative methods of analysis: Multitrait-multimethod data. *Educational and Psychological Measurement*, **30**, 833-853.
- Brennan, R. L. (1992). *Elements of generalizability theory. Revised ed.* Iowa City, IA: American College Testing Program.
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer.
- Browne, M. W. (1984). The decomposition of multitrait-multimethod matrices. *British Journal of Mathematical and Statistical Psychology*, **37**, 1-21.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, **56**, 81-105.
- Campbell, D. T., & O'Connell, E. J. (1967). Methods factors in multitrait-multimethod matrices: Multiplicative rather than additive? *Multivariate Behavioral Research*, **2**(4), 409-426.
- Carroll, J. B. (1952). Ratings on traits measured by a factored personality inventory. *Journal of Abnormal and Social Psychology*, **47**, 626-632.
- Conway, J. M. (1996). Analysis and design of multitrait-multirater performance appraisal studies. *Journal of Management*, **22**, 139-162.
- Conway, J. M. & Huffcutt, A. I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance*, **10**, 331-360.
- Conway, J. M., Lombardo, K., & Sanders, K. C. (2001). A meta-analysis of incremental validity and nomological networks for subordinate and peer ratings. *Human Performance*, **14**, 267-303.
- Coenders, G., & Saris, W. E. (2000). Testing nested additive, multiplicative, and general multitrait-multimethod models. *Structural Equation Modeling*, **7**, 219-250.

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, **16**(3), 297-334.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.) *Educational measurement*. 2nd ed. pp. 443-507. Washington, DC: American Council on Education.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability of scores and profiles*. New York: Wiley.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, **51**, 281-302.
- Cudeck, R. (1988). Multiplicative models and MTMM matrices. *Journal of Educational Statistics*, **13**(2), 131-147.
- Dempster A.P., Laird N.M., & Rubin D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)* **39**(1), 1-38.
- Duane, S., Kennedy, A. D., Pendleton, B. J. & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, **195**(2), 216-222.
- Edwards, M. R., & Ewen, I. J. (1996a). *360° feedback: The powerful new model for employee assessment & performance improvement*. New York: AMACOM American Management Association.
- Edwards, M. R., & Ewen, I. J. (1996b). How to manage performance and pay with 360-degree feedback. *Comparison & Benefits Review*, **28**, 41-46.
- Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, **65**(2), 241-261.
- Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model. *Psychological Methods*, **8**(1), 38-60.
- Eid, M., Nussbeck, F. W., Geiser, C., Cole, D., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*, **13**(3), 230-253.

- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In Linn, R. L. (Ed.), *Educational measurement*. 3rd ed. New York: McMillan. (リン R. L. (編) 池田央・藤田恵璽・柳井晴夫・繁榊算男 (監訳) (1992). 信頼性. 教育測定学 上. C.S.L. 学習評価研究所 pp.147-209.)
- Forsythe, G. B., McGaghie, W. C. & Friedman, C. P. (1986). Construct validity of medical clinical competence measures: A multitrait-multimethod matrix study using confirmatory factor analysis. *American Educational Research Journal*, **23**(2), 315-336.
- 学校法人産業能率大学総合研究所. ビジネス基礎力診断 S-BASE(エス・ベース) <http://www.hj.sanno.ac.jp/cp/page/7638> (2014年2月).
- 学校法人産業能率大学総合研究所. プロ人材特性診断 S-Pro(エスプロ)II <http://www.hj.sanno.ac.jp/cp/page/5714> (2014年2月).
- 学校法人産業能率大学総合研究所. マネジャー力診断 S-ATOM(エスアトム) <http://www.hj.sanno.ac.jp/cp/page/5716> (2014年2月).
- 学校法人産業能率大学総合研究所. (2012). -最適人事と能力開発のための-アセスメントツール総合ガイド 2012-2013.
- Gelman, A. (1996). Inference and monitoring convergence. In W. R. Gilks, S. Richardson & D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*. pp.131-143. London: Chapman and Hall.
- Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, **7**, 457-511.
- Goffin, R. D., & Jackson, D. N. (1992). Analysis of multitrait-multirater performance appraisal data: Composit direct product method versus confirmatory factor analysis. *Multivariate Behavioral Research*, **27**, 363-385.
- Grayson, D., & Marsh, H. W. (1994). Identification with deficient rank loading matrices in confirmatory factor analysis: multitrait-multimethod matrix. *Psychometrika*, **59**, 121-134.
- Greguras, G. J., & Robie, C. (1998). A new look at within-source reliability of 360-degree feedback ratings. *Journal of Applied Psychology*, **83**, 960-968.
- 南風原朝和 (2001). 量的調査—尺度の作成と相関分析. 南風原朝和, 市川伸一, 下山晴彦 (編) 心理学研究法入門—調査・実験から実践まで— 東京大学出版 pp.63-91.

- 南風原朝和 (2002). 心理統計学の基礎-統合的理解のために-. 有斐閣.
- 南風原朝和 (2011). 臨床心理学をまなぶ7 量的研究法. 東京大学出版会.
- Harris, M. M. & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor rating. *Personnel Psychology*, **41**(1), 43-62.
- Hastings, W. K. (1970). Monte carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**(1), 97-109.
- 平井洋子 (2006). 測定の妥当性からみた尺度構成-得点の解釈を保證できますか. 吉田寿夫(編著)心理学研究法の新しいかたち 誠信書房.
- 池田央 (1992a). 行動計量シリーズ7 現代テスト理論. 朝倉書店.
- 池田央 (1992b). 妥当性. リン R. L. (編)池田央・藤田恵璽・柳井晴夫・繁榊算男(監訳). 教育測定学上. C.S.L. 学習評価研究所 pp.19-145.
- 入江崇介・鷺坂由紀子・舛田博之・二村英幸 (2004). 多面観察評価手法の特性に関する検討-一般化可能性理論によるアプローチ-. 日本テスト学会第2回大会発表論文抄録集, 68-69.
- 入江崇介・鷺坂由紀子・舛田博之・二村英幸 (2005). 多面観察評価における上司・同僚・部下の評定結果の統合について-多変量一般化可能性理論による一般化可能性の検討-. 日本テスト学会第3回大会発表論文抄録集, 60-63.
- Jackson, D. N. (1969). Multimethod factor analysis in the evaluation of convergent and discriminant validity. *Psychological Bulletin*, **72**(1), 30-49.
- Jackson, D. N. (1975). Multimethod factor analysis: A reformulation. *Multivariate Behavioral Research*, **10**, 259-276.
- Kalleberg, A. L., & Kluegel, J. R. (1975). Analysis of the multitrait-multimethod matrix: Some limitations and an alternative. *Journal of Applied Psychology*, **60**(1), 1-9.
- Kavanagh, M. J., MacKinney, A. C., & Wolins, L. (1971). Issues in managerial performance: Multitrait-multimethod analysis of ratings. *Psychological Bulletin*, **75**(1), 34-49.
- Kelley, T. L. (1927). *Interpretation of educational measurements*. New York: Macmillan p.14.
- Kelly, E. L., & Fiske, D. W. (1951). *The prediction of performance in clinical psychology*, Ann Arbor, MI: University of Michigan Press.

- Kenny, D. A. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology*, **12**, 247-252.
- Kenny, D. A. (1994). The multitrait-multimethod matrix: Design, analysis and Conceptual Issues. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and Theory*, pp. 111-124. Hillsdale, NJ: Lawrence Erlbaum.
- Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, **112**(1), 165-172.
- Kothandapani, V. (1971). Validation of feeling, belief, and intention to act as three components of attitude and their contribution to the prediction of contraceptive behavior. *Journal of Personality and Social Psychology*, **19**, 321-333.
- Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods*, **7**(2), 228-244.
- Landy, F. J. & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, **87**(1), 72-107.
- Lawler, E. E, III. (1967). The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology*, **51**, 369-381.
- Lehmann, D. R. (1988). An alternative Procedure for assessing convergent and discriminant validity. *Applied Psychological Measurement*, **12**(4), 411-423.
- Lepsinger, R. & Lucia, A. D. (2009). *The art and science of 360-degree feedback. 2nd Ed.* SanFrancisco: Jossey-Bass A Wiley Import.
- Levin, J. (1988). Multiple group factor analysis of multitrait-multimethod matrices. *Multivariate Behavioral Research*, **23**, 469-479.
- Levin, J. (1973). Bifactor analysis of a multitrait-multimethod matrix of leadership criteria in small groups. *Journal of Social Psychology*, **89**, 295-299.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.
- Lomax, R. G., & Algina, J. (1979). Comparison of two procedures for analyzing multitrait multimethod matrices. *Journal of Educational Measurement*, **16**(3), 177-186.

- Maas, C. J. M., Lensvelt-Mulders, G. J. L. M. & Hox, J. J. (2009). A multilevel multitrait-multimethod analysis. *Methodology*, **5**, 72-77.
- Mabe, P. A. & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, **67**(3), 280-296.
- Manson, T. M., Levine, E. L., & Brannick, M. T. (2000). The construct validity of task inventory ratings: A multitrait-multimethod analysis. *Human Performance*, **13**, 1-22.
- Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod data: many problems and a few solutions. *Applied Psychological Measurement*, **13**(4), 335-361.
- Marsh, H. W., & Bailey, M. (1991) Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement*, **15**(1), 47-70.
- Marsh, H. W., & Grayson, D. (1995). Latant variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling. Concepts, issue, and applications*. Thousands Oals, CA: Sage. pp.177-198.
- 松本隆 (2004). ハミルトニアン・モンテカルロ法によるベイズ的学習と予測. 松本隆, 石黒真木夫, 乾敏郎, 田邊國士 (著) 階層ベイズモデルとその周辺- 時系列・画像・認知への応用- 岩波書店. pp.145-157.
- McDonald, R. P. (1978). Generalizability in factorable domains: "Domain validity and generalizability". *Educational and Psychological Measurement*, **38**(1), 75-79.
- McDonald, R. P. (1999) *Test theory: A unified treatment*. New Jersey: Lawrence Erlbaum.
- Messick, S. (1989). Validity. In Linn, R. L. (Ed.), *Educational measurement*. 3rd ed. pp. 13-103. New York: McMillan. (リン R. L. (編) 池田央・藤田恵璽・柳井晴夫・繁榊算男 (監訳) (1992). 妥当性. 教育測定学 上. C.S.L. 学習評価研究所 pp.19-145.)
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, **50**, 741-749.

- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**(6), 1087-1092.
- Millsap, R. R. (1992). Sufficient conditions for rotational uniqueness in the additive MTMM model. *British Journal of Mathematical and Statistical Psychology*, **45**, 125-138.
- 森敏昭 (2009). 心をどう測るのか—測定の妥当性と信頼性—. 無藤隆・森敏昭・池上知子・福丸由佳 (編) やわらかアカデミズム・<わかる>シリーズ よくわかる心理学. ミネルヴァ書房, pp.14-15.
- 森敏昭・吉田寿夫 (1990). 心理学のためのデータ解析テクニカルブック. 北大路書房.
- Mount, K. M. (1984). Psychometric properties of subordinate rating of managerial performance. *Personnel Psychology*, **37**, 687-702.
- Mount, M. K., Judge, T. A, Scullen, S. E. Sytsma, M. R. & Hezlett, S. A. (1998). Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology*, **51**, 557-576.
- Muis, K. R., Winne, P. H., & Jamieson-Noel, D. (2007). Using a multitrait-multimethod analysis to examine conceptual similarities of three self-regulated learning inventories. *British Journal of Educational Psychology*, **77**, 177-195.
- Neal, R. M. (1996). *Bayesian Learning for Neural Networks*. Lecture Notes in Statistics 118, NY: Springer-Verlag.
- Neal, R. M. (2011). MCMC Using Hamiltonian Dynamics. In *Handbook of Markov Chain Monte Carlo*. Brooks, S., Gelman, A., Galin L. Jones G. L. & Meng, X-L., (Eds.). Chapman & Hall/CRC, pp.113-162.
- 中妻照雄 (2007) 入門ベイズ統計学 朝倉書店.
- 日本経営協会総合研究所. 多面観察<マネジメント版・リーダーシップ版> http://www.noma.co.jp/service/s_as/sas01_01.html (2014年2月).
- 日本能率協会マネジメントセンター. NAVI360 http://www.jmam.co.jp/product/service/assess/item_navi360.html (2014年2月).
- 日本能率協会マネジメントセンター. RoundReview http://www.jmam.co.jp/product/service/assess/item_roundreview.html (2014年2月).

- 二村英幸 (1998). 人事アセスメントの科学 産能大学出版部.
- 二村英幸 (2001). 人事アセスメント入門 日本経済新聞社.
- Ostrom, T. M. (1969). The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Experimental Social Psychology*, **5**, 12-30.
- 大沢武志 (2000). 経営人事における人事アセスメント. 大沢武志・芝祐順・二村英幸 (編) 人事アセスメントハンドブック 金子書房, pp.3-22.
- リクルートマネジメントソリューションズ. MOA|360° 評価システム http://www.recruit-ms.co.jp/service/service_detail/org_key/A005/ (2014年2月).
- Roshal, S. M., Frieze, I., & Wood, J. T. (1971). A multitrait-multimethod validation of measures of student attitudes toward school, toward learning, and toward technology in sixth grade children. *Educational and Psychological Measurement*, **31**, 999-1006.
- Schmitt, N. (1978). Path analysis of multitrait-multimethod matrices. *Applied Psychological Measurement*, **2**(2), 157-173.
- Schmitt, N., Coyle, B. W., & Saari, B. B. (1977). A review and critique of analyses of multitrait-multimethod matrices. *Multivariate Behavioral Research*, **12**, 447-478.
- Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement*, **10**(1), 1-22.
- 関本昌秀 (1965). わが国における人事考課の反省—管理者層の人事考課を中心として 慶應ビジネスフォーラム.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. London: Sage Publications.
- Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning in life questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology*, **53**, 80-93.
- 高橋潔 (2001). 多面評価法 (360度フィードバック法) に関する多特性多評価者行列分析. *経営行動科学* **14**(2), 67-85.
- 高橋潔 (2010). 人事評価の総合科学—努力と能力と行動の評価 白桃書房.

- 田中堅一郎(編) (2011) 産業・組織心理学エッセンシャルズ【改訂三版】 ナカニシヤ出版.
- 豊田秀樹 (2000). 共分散構造分析 [応用編] 朝倉書店.
- 豊田秀樹 (2008). マルコフ連鎖モンテカルロ法 朝倉書店.
- Thornton, G. C. (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology*, **33**(2), 263-271.
- 上田敏晶 (1974). 若手のエネルギーを引き出す“人物評価”: 神戸製鋼所の『多面観察調査』法. *マネジメント* **33**(1), 107-112.
- 渡部洋 (1999). ベイズ統計学入門 福村出版.
- Werts, C. E., & Linn, R. L. (1970). Path analysis: Psychological examples. *Psychological Bulletin*, **74**, 193-212.
- Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, **9**(1), 1-526.
- Wothke, W., & Browne, M. W. (1990). The direct product model for the MTMM matrix parameterized as a second order factor analysis model. *Psychometrika*, **55**(2), 255-262.
- Wothke, W. (1984) The estimation of trait and method components in multitrait-multimethod data. Unpublished doctoral dissertation. University of Chicago.